UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

**Explorando Internet das Coisas e Inteligência Artificial no
contexto de Saúde em Casas Inteligentes: uma abordagem
física e emocional**

Leandro Yukio Mano Alves

Tese de Doutorado do Programa de Pós-Graduação em Ciências de
Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: _____

Leandro Yukio Mano Alves

Explorando Internet das Coisas e Inteligência Artificial no contexto de Saúde em Casas Inteligentes: uma abordagem física e emocional

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientador: Prof. Dr. Jó Ueyama

USP – São Carlos
Outubro de 2019

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M285e Mano Alves, Leandro Yukio
Explorando Internet das Coisas e Inteligência
Artificial no contexto de Saúde em Casas
Inteligentes: uma abordagem física e emocional /
Leandro Yukio Mano Alves; orientador Jó Ueyama. --
São Carlos, 2019.
123 p.

Tese (Doutorado - Programa de Pós-Graduação em
Ciências de Computação e Matemática Computacional) --
Instituto de Ciências Matemáticas e de Computação,
Universidade de São Paulo, 2019.

1. Internet das Coisas. 2. Saúde em Casas
Inteligentes. 3. Aspectos físicos e emocionais. 4.
Abordagem multimodal. 5. Arquitetura baseada em
camadas. I. Ueyama, Jó, orient. II. Título.

Leandro Yukio Mano Alves

**Exploiting the Internet of Things with Artificial Intelligence in
the context of Health Smart Homes: a physical and
emotional approach**

Doctoral dissertation submitted to the Institute of
Mathematics and Computer Sciences – ICMC-USP, in
partial fulfillment of the requirements for the degree of
the Doctorate Program in Computer Science and
Computational Mathematics. *FINAL VERSION*

Concentration Area: Computer Science and
Computational Mathematics

Advisor: Prof. Dr. Jó Ueyama

USP – São Carlos
October 2019

Este trabalho é dedicado aos meus pais, Cleudo e Elza, pelo sacrifício ilimitado em todos os sentidos, que mais do que me proporcionar uma boa infância, formaram os fundamentos do meu caráter e me apontaram o caminho a seguir. Obrigado por serem a minha referência de tantas maneiras e estarem sempre presentes na minha vida de uma forma indispensável.

A minha irmã Camila e família, que apesar de tantas provações esteve sempre disposta a me estender a mão e me oferecer ajuda, carinho, companhia. Que eu possa servir de exemplo para Livia e Mariana (sobrinhas), para que elas possam seguir um caminho de luta e trabalho para conquistas de seus sonhos.

Aos meus familiares, que me apoiaram de todas as formas possíveis, incentivando, aconselhando em toda minha vida.

Aos amigos de perto e de longe, pelo amor e preocupação demonstrados através de ligações, visitas e e-mails. Obrigado, vocês que aliviaram minhas horas difíceis, me alimentando de certezas, força e alegria.

Em especial, dedico a minha esposa Manuela, pela paciência, pelo apoio incondicional e constante ensinamento em todos os momentos, me proporcionando experiências, segurança, conforto e alegrias que são importantes para que eu possa me tornar uma pessoa melhor. Muito obrigado, nunca será suficiente para demonstrar a grandeza do que recebi de todos.

AGRADECIMENTOS

Ao Profº. Dr. Jó Ueyama, professor e orientador pelo desprendimento ao escolher me dar apoio, muito obrigado pela dedicação e ajuda durante essa etapa da minha vida.

Agradeço a Universidade de São Paulo - USP e ao Laboratório de Sistemas Interativos Web e de Sistemas Multimídia - Intermídia pelo incentivo, apoio e infraestrutura para a pós-graduação e o desenvolvimento deste projeto. Um agradecimento especial aos integrantes do *Wireless Networks Group - WinesGroup* pela disponibilidade, esforço e dedicação para com os membros do grupo, em especial a minha pessoa.

Agradeço a Universidade de Gronigen, Holanda, pelo apoio e infraestrutura para o desenvolvimento de parte deste projeto e, em especial, ao Prof. Dr. Michael Biehl pelo orientação e apoio durante minha permanência na Universidade de Groningen para o estágio no exterior.

Agradeço ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) pelo apoio no início do curso de pós-graduação, por disponibilizar uma bolsa de estudo destinado a alunos de doutorado - **processo nº 142390/2016-0, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).**

Agradeço a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por apoiar este discente no restante e maior parte do curso de pós-graduação, por disponibilizar uma bolsa de estudo destinado a alunos de doutorado - **processo nº 2016/14267-7, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).**

Agradeço novamente a Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) por apoiar este discente com Bolsa Estágio de Pesquisa no Exterior (BEPE), por disponibilizar uma bolsa de estudo destinado a alunos a bolsistas de Doutorado da FAPESP. - **processo nº 2017/21054-2, Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP).**

*“A educação é a arma mais poderosa
pela qual se pode mudar o mundo.”
(NELSON MANDELA)*

RESUMO

MANO, L. Y. **Explorando Internet das Coisas e Inteligência Artificial no contexto de Saúde em Casas Inteligentes: uma abordagem física e emocional**. 2019. 123 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Devido ao aumento da população idosa ou com limitações físicas/mentais é possível observar o crescimento de estudos na área de Internet das Coisas com o objetivo de monitorar a saúde e auxiliar no gerenciamento e melhora na qualidade de vida dessa parte da população. Nesse sentido, a abordagem baseada em Internet das Coisas aplicada em ambientes médicos e casas inteligentes tem o objetivo de fornecer conectividade entre o paciente e o ambiente ao seu redor, provendo mecanismos para ajudar em diagnósticos e prevenção de acidentes e/ou doenças. Nesse contexto surge a oportunidade de explorar sistemas computacionais para identificar o estado físico e emocional, em tempo real, de indivíduos com alguma limitação para monitorar a saúde; por exemplo, identificar o comportamento da rotina do usuário e emitir alertas aos familiares e/ou equipe médica sobre algum evento anormal ou identificar indícios de distúrbios emocionais. Ainda, com base na Inteligência Artificial é possível que sistemas computacionais possam “aprender” e se adaptar ao contexto que se encontra, por exemplo aprender e se adaptar a quantidade de exercícios e/ou estado emocional do usuário em determinadas situações, combinando conceitos tanto de *Internet of Things* quanto de Inteligência Artificial. Assim, este projeto tem como objetivo desenvolver e avaliar um modelo que possa: i.) identificar o estado físico e emocional do usuário; ii.) prover um mecanismo que possa monitorar de maneira inteligente as atividades do cotidiano do usuário e; iii.) explorar a abordagem de integração de dados com a utilização de múltiplos sensores IoT para uma melhor interação entre dispositivos computacionais no ambiente *Health Smart Homes*.

Palavras-chave: Internet das Coisas, Saúde em Casas Inteligentes, Aspectos físicos e emocionais, Abordagem multimodal, Arquitetura baseada em camadas.

ABSTRACT

MANO, L. Y. **Exploiting the Internet of Things with Artificial Intelligence in the context of Health Smart Homes: a physical and emotional approach.** 2019. 123 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2019.

Due to the increase in the elderly population or with physical/mental limitations it is possible to observe the growth of studies in the field of Internet of Things with the objective of monitoring health and assisting in the management and improvement of the quality of life of this part of the population . In this sense, the Internet of Things approach applied in medical environments and smart homes aims to provide connectivity between the patient and the environment around them, providing mechanisms to aid in diagnosis and prevention of accidents and/or diseases. In this context the opportunity arises to explore computational systems to identify the physical and emotional state, in real time, of individuals with some limitation to monitor health; for example, identifying the behavior of the user's routine and issuing alerts to family members and/or medical staff about some abnormal event or identifying evidence of emotional disturbances. Also, based on Artificial Intelligence, it is possible for computational systems to “learn” and adapt to the context they are in, for example learning and adapting the amount of exercises and/or emotional state of the user in certain situations, combining concepts both of Internet of Things and of Artificial Intelligence. Thus, this project aims to develop and evaluate a model that can: i) identify the physical and emotional state of the user; ii) provide a mechanism that can intelligently monitor the daily activities of the user and; iii) explore the data integration approach with the use of multiple IoT sensors for better interaction between computing devices in the Health Smart Homes environment.

Keywords: Internet of Things, Health Smart Homes, Physical and emotional aspects, Multi-modal approach, Layer-based Architecture.

LISTA DE ILUSTRAÇÕES

Figura 1 – Modelo Circumplexo das emoções.	32
Figura 2 – Emoções básicas propostas por Ekman (1973) e Ekman (2006), com a expressão Neutro.	33
Figura 3 – Cenário de classificação multimodal da emoção considerando abordagem multimodal no nível de característica.	37
Figura 4 – Processo de mapeamento da face.	53
Figura 5 – Boxplots das acurácias apresentadas pelos classificadores para a identificação do indivíduo.	54
Figura 6 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção pela face.	56
Figura 7 – Matriz de Confusão da classificação da face para emoção (abordagem SVM).	58
Figura 8 – Representação dos componentes de frequência da fala.	58
Figura 9 – Espectrograma da frase “Gli operari si alzarò presto” pronunciada de maneira negativa e positiva no idioma italiano.	59
Figura 10 – Representação do <i>pipe</i> de extração de <i>features</i> da voz.	60
Figura 11 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção do espectro de energia da voz.	61
Figura 12 – Matriz de Confusão da classificação da voz para emoção (abordagem kNN).	62
Figura 13 – Etapas para o processamento da descrição da fala.	64
Figura 14 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção mediante a descrição da fala.	65
Figura 15 – Matriz de Confusão da classificação da voz da descrição da fala (abordagem N. Bayes).	66
Figura 16 – Estrutura para a classificação multimodal da emoção baseada na face, tom de voz e descrição da fala.	68
Figura 17 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção dos vídeos.	69
Figura 18 – Matriz de Confusão referente a classificação da face (SVM), voz (kNN) e contexto (N. Bayes).	70
Figura 19 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção dos vídeos, levando em consideração a abordagem multimodal.	71
Figura 20 – Matriz de Confusão da classificação multimodal da face, voz e contexto (abordagem SVM).	72

Figura 21 – Simulação clínica - “Cuidados de enfermagem ao paciente com ascite e desconforto respiratório seguido de vômito”. Imagem dos estudantes submetidos ao experimento no cenário simulado.	74
Figura 22 – Número de emoções identificadas por momentos na simulação clínica. . . .	75
Figura 23 – Frequência de emoções segundo o Modelo Circumplexo (SCHERER, 2005). . .	76
Figura 24 – Exemplo de aplicação de monitoramento de quedas.	80
Figura 25 – Estrutura para a classificação de atividades do usuário, com ênfase em uma possível queda.	81
Figura 26 – Boxplots das acurácias apresentadas pelos classificadores para determinar as atividades do usuário.	82
Figura 27 – Matriz de Confusão da classificação de atividades (abordagem do SVM). . .	83
Figura 28 – Contexto não intrusivo para o monitoramento do ambiente para a qualidade do sono.	85
Figura 29 – Boxplots referente a sensibilidade e qualidade dos sensores de luminosidade, ruído e movimento.	87
Figura 30 – Testes do monitoramento dos cenários por um período prolongado.	89
Figura 31 – Cenário do ENLACE baseado em três níveis de processamento em tempo de execução.	92
Figura 32 – Protótipo do dispositivo de monitoramento residencial.	93
Figura 33 – Fluxo do processo de comunicação entre as camadas do ENLACE para a classificação de aspectos físicos e emocionais do usuário.	94
Figura 34 – Número de solicitações e tempo decorrido dos serviços no nível do dispositivo local.	97
Figura 35 – Número de solicitações e tempo decorrido dos serviços no nível do servidor local.	99
Figura 36 – Número de solicitações e tempo decorrido dos serviços no nível do servidor em nuvem.	100

LISTA DE QUADROS

Quadro 1 – Análise dos trabalhos identificados na literatura referente a classificação de aspectos emocionais.	42
Quadro 2 – Análise dos trabalhos identificados na literatura referente a classificação de aspectos físicos.	45
Quadro 3 – Análise dos trabalhos identificados na literatura referente ambientes HSH. .	48
Quadro 4 – Descrição das bases de dados para a classificação emocional da face. . . .	56
Quadro 5 – Descrição das bases de dados para a classificação emocional da voz. . . .	60
Quadro 6 – Descrição do balanceamento da base de dados OMG-Emotion Dataset. . .	68
Quadro 7 – Cenários do experimento para o monitoramento do ambiente.	86
Quadro 8 – Classificação do ambiente para o repouso.	88
Quadro 9 – Configurações do ambiente do ENLACE.	96
Quadro 10 – Configurações do experimento - taxa de requisições.	97

LISTA DE TABELAS

Tabela 1 – Média (%) das acurácias e os p -valores dos conjuntos de resultados para a identificação de indivíduo.	54
Tabela 2 – P -valores da comparação de pares realizada com o teste <i>Wilcoxon Rank Sum</i> para a identificação do indivíduo.	55
Tabela 3 – Média (%) das acurácias e os p -valores dos conjuntos de resultados para a classificação da emoção.	57
Tabela 4 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela face.	57
Tabela 5 – Média (%) das acurácias e os p -valores dos conjuntos de resultados da classificação da voz.	61
Tabela 6 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela voz.	62
Tabela 7 – Média (%) das acurácias e os p -valores dos conjuntos de resultados da classificação da descrição da fala.	65
Tabela 8 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela descrição da fala.	66
Tabela 9 – Acurácia média (%) dos algoritmos avaliados.	69
Tabela 10 – Acurácia média (%) e p -valores (<i>Shapiro Wilk</i>) dos algoritmos avaliados. . .	71
Tabela 11 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela abordagem multimodal.	71
Tabela 12 – Relação dos momentos da simulação e emoções (%).	75
Tabela 13 – Acurácia média (%) e p -valores (<i>Shapiro Wilk</i>) dos algoritmos avaliados. . .	83
Tabela 14 – Comparações pareadas utilizando testes- T na classificação de atividades. . .	83
Tabela 15 – Relação da faixa de intensidade dos valores dos sensores do monitoramento do ambiente.	88

LISTA DE ABREVIATURAS E SIGLAS

CK+	<i>Extended Cohn-Kanade</i>
CNN	<i>Convolutional Neural Network</i>
ENLACE	<i>EmotioNal Level Architecture Communication hEalthcare</i>
FACS	<i>Facial Action Coding System</i>
HSH	<i>Health Smart Homes</i>
IBM	<i>International Business Machines</i>
IHC	Interação Humano-Computador
IoT	<i>Internet of Things</i>
IoTDSM	<i>Internet of Things Data as a Service Module</i>
kNN	k-Nearest Neighbor classification
LEDs	<i>Light Emitting Diode</i>
LIWC	<i>Linguistic Inquiry and Word Count</i>
MFCC	<i>Mel Frequency Cepstral Coefficient</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
N. Bayes	Naive Bayes
NLTK	<i>Natural Language Toolkit</i>
OMG-Emotion Dataset	One-Minute-Gradual Emotion Dataset
PLN	Processamento de Linguagem Natural
RaFD	<i>Radboud Faces</i>
SVM	Support Vector Machine
Árv. Dec.	Árvore de Decisão

SUMÁRIO

1	INTRODUÇÃO	25
1.1	Objetivos	27
1.2	Estrutura da Tese	27
2	FUNDAMENTAÇÃO TEÓRICA	29
2.1	Internet das Coisas aplicada na Saúde em Casas Inteligentes	29
2.2	Conhecimento prévio sobre emoções - Teoria Componencial das Emoções	31
2.3	Inteligência Artificial para a classificação do estado físico e emocional	34
2.4	Considerações finais	37
3	TRABALHOS RELACIONADOS	39
3.1	Classificação de aspectos emocionais	39
3.2	Classificação de aspectos físicos	42
3.3	Arquitetura para Saúde em Casas Inteligentes	45
3.4	Considerações finais	48
4	INTELIGÊNCIA COMPUTACIONAL PARA IDENTIFICAÇÃO DE ASPECTOS EMOCIONAIS	51
4.1	Identificação do usuário e classificação da emoção pela face	51
4.1.1	<i>Método para a análise facial - Características geométricas</i>	52
4.1.2	<i>Reconhecimento do indivíduo - Avaliação e discussão</i>	53
4.1.3	<i>Classificação da emoção - Avaliação e discussão</i>	55
4.2	Classificação da emoção pelo tom de voz	58
4.2.1	<i>Método para análise da voz - Espectro de energia</i>	59
4.2.2	<i>Avaliação e discussão</i>	60
4.3	Classificação da emoção pela descrição da fala	63
4.3.1	<i>Método para análise da descrição da fala</i>	63
4.3.2	<i>Avaliação e discussão</i>	64
4.4	Abordagem multimodal para a classificação da emoção	66
4.4.1	<i>Metodologia multimodal para a emoção</i>	67
4.4.2	<i>Avaliação e discussão</i>	67
4.5	Classificação multimodal da emoção em ambiente real - Simulação clínica	72

4.5.1	<i>Metodologia na classificação emocional em simulação clínica</i>	73
4.5.2	<i>Avaliação e discussão</i>	75
4.6	<i>Considerações finais</i>	77
5	INTELIGÊNCIA COMPUTACIONAL PARA IDENTIFICAÇÃO DE ASPECTOS FÍSICOS	79
5.1	<i>Identificação de atividades</i>	79
5.1.1	<i>Inteligência Artificial para detecção de quedas</i>	80
5.1.2	<i>Avaliação e discussão na detecção de atividades</i>	81
5.2	<i>Qualidade do ambiente para o repouso</i>	84
5.2.1	<i>Sistema de monitoramento do ambiente - Qualidade do ambiente para o repouso</i>	85
5.2.2	<i>Metodologia para o monitoramento do ambiente</i>	86
5.2.3	<i>Avaliação e discussão dos resultados</i>	87
5.3	<i>Considerações finais</i>	89
6	ARQUITETURA PARA O MONITORAMENTO DE ASPECTOS FÍSICOS E EMOCIONAIS PARA SAÚDE EM CASAS INTELIGENTES	91
6.1	<i>ENLACE - Arquitetura baseada em camadas para o monitoramento em ambientes residenciais</i>	91
6.1.1	<i>Protótipo para o monitoramento de aspectos físicos e emocionais em ambientes residenciais</i>	92
6.1.2	<i>Monitoramento físico e emocional como serviço</i>	94
6.1.3	<i>Tomada de decisão de distribuição de carga entre camadas</i>	95
6.1.4	<i>Avaliação e discussão</i>	96
6.1.4.1	<i>Configuração do experimento</i>	96
6.1.4.2	<i>Resultados experimentais e discussão</i>	97
6.2	<i>Considerações finais</i>	101
7	CONCLUSÃO	103
7.1	<i>Aplicabilidade e trabalhos futuros</i>	104
7.2	<i>Principais contribuições</i>	104
	REFERÊNCIAS	109
ANEXO A	COMITÊ DE ÉTICA Nº 155/2013	123

INTRODUÇÃO

O número de pessoas com alguma deficiência, mobilidade reduzida e idosos é cada vez maior em vários países, principalmente nos países desenvolvidos (GONÇALVES *et al.*, 2013b; MANO *et al.*, 2016b), e cada vez mais essa parte da população procura morar sozinha, em busca de sua independência e autonomia (MANO *et al.*, 2016b). Porém, muitas vezes, possuem maiores dificuldades em seu cotidiano e são mais suscetíveis a acidentes e doenças do que o restante da população. Isso, muitas vezes, leva ao uso da tecnologia para monitorar seu comportamento, cotidiano e/ou enquanto se recuperam em casa (ROMERO *et al.*, 2009).

A tecnologia nesse contexto é viável, principalmente em países onde o número de pessoas nessas condições é alto, podendo fornecer sistemas que possam monitorá-los e emitir alertas para a equipe médica e/ou familiares sempre que necessário, tornando o uso de tecnologias que possibilitem monitorar a saúde das pessoas extremamente relevante (ROMERO *et al.*, 2009). Nesse sentido, o conceito de Saúde em Casas Inteligentes (*Health Smart Homes* (HSH)) emerge da combinação de Telemedicina, Internet das Coisas (*Internet of Things* (IoT)) e Sistemas de Informação; sendo definido como uma casa inteligente, equipada com dispositivos especializados para o cuidado da saúde à distância, incluindo smartphones, tecnologias embarcadas e vestíveis (por exemplo, *smartwatch*) (MANO *et al.*, 2016a; MANO *et al.*, 2016b). Este contexto tem sido largamente explorado em pesquisas na literatura (RIALLE *et al.*, 2002; STANKOVIC *et al.*, 2005; AUGUSTO *et al.*, 2007; ROMERO *et al.*, 2009; DOHR *et al.*, 2010; MANO *et al.*, 2016a; Mano, 2018), onde os autores propõem soluções tecnológicas para auxiliar no monitoramento de pessoas com necessidades. O objetivo é fornecer subsídios às pessoas com alguma dificuldade ou doenças, para que possam viver suas vidas de uma forma menos dependente.

Diante desse panorama, observa-se que a maioria dos esforços estão concentrados na melhoria da vida diária dos usuários, fornecendo-lhes *gadgets*¹ específicos, como alarmes de

¹ *Gadget* é um equipamento que tem um propósito e uma função específica, prática e útil no cotidiano. São comumente chamados de *gadgets* dispositivos eletrônicos portáteis como celulares, smartphones, leitores de MP3, sensores, entre outros.

alerta (MANO *et al.*, 2016b). Por exemplo, a Intel Corporation desenvolveu um sistema baseado em ultrassom que identifica quedas e movimentos irregulares dos usuários, como parte de seu sistema de IoT (INTEL, 2016). Romero *et al.* (2009) descrevem sobre o uso de sensores corporais e dispositivos específicos quando prestam assistência a idosos e pessoas com necessidades. Augusto *et al.* (2007) faz o uso de câmeras para melhorar a análise do ambiente (p. ex., detecção de possíveis situações de perigo) proporcionando um sistema não intrusivo. No entanto, as abordagens (baseado em sensores e câmera) tem claramente vantagens e desvantagens, e devem ser aplicados com objetivos específicos, o que pode limitar sua utilização, como por exemplo, tais sistema não podem reconhecer qualquer tipo de aspecto comportamental e/ou emocional, que também são cruciais durante o cuidado domiciliar.

No setor comercial, algumas iniciativas têm gerado dispositivos para interação com o usuário e monitoramento de ambientes residenciais. Por exemplo, o *Google Home* (GOOGLE, 2018), desenvolvido pelo Google permite que os usuários falem comandos de voz para interagir com serviços, internos e de terceiros, permitindo aos usuários ouvir música, controlar a reprodução de vídeos ou fotos, receber atualizações de notícias totalmente por voz por meio do assistente pessoal. Outra iniciativa, desenvolvido pela *International Business Machines* (IBM) (IBM, 2018b) tem iniciado ações voltadas para a área da saúde e oferece ferramentas para propiciar a predição baseada em dados, gerados por tecnologias móveis e sensores, e criam quantidades muito grandes de dados que podem ser incorporados ao processo de monitoramento residencial. No entanto, ambas as iniciativas, apesar de fornecer diversas funcionalidades para casas inteligentes, não abordam nenhum aspecto relacionado a saúde emocional e física do usuário, como é o caso do *Google Home*, ou as soluções oferecidas são voltadas para empresas e investidores e não ao usuário final, caso da IBM.

É importante ressaltar que aspectos emocionais desempenham um papel importante no cotidiano e na recuperação de pessoas com alguma necessidade específica (FREDRICKSON; LEVENSON, 1998; OSTIR *et al.*, 2008; MANO *et al.*, 2016a). Estudos abordam a saúde emocional do indivíduo como um sintoma importante para o diagnóstico de várias doenças (SALEM; KRING; KERR, 1996; MANDAL; PANDEY; PRASAD, 1998; GUR *et al.*, 2006; IVANOVA, 2012), tais como: a esquizofrenia, depressão, autismo e transtorno bipolar. Ressalta-se que, as doenças mencionadas podem ser causadas pelo tempo excessivo em emoções negativas, falta de expressões emocionais ou a instabilidade das emoções expressadas (SALEM; KRING; KERR, 1996; MANDAL; PANDEY; PRASAD, 1998; GUR *et al.*, 2006). Além do uso da saúde emocional para diagnosticar doenças, estados emocionais podem influenciar a interação social e podem ser utilizados para intervir na realização de uma tarefa, ou até mesmo para tomar decisões em prol do usuário; por exemplo, sugerindo um gênero de filme para aliviar o estresse com base na sua emoção atual (GONÇALVES *et al.*, 2013a; MANO *et al.*, 2016a).

Um dos grandes desafios dos pesquisadores é encontrar formas de prover sistemas capazes de reconhecer, interpretar e reagir de modo inteligente e sensível, que possam atender

os requisitos do maior número possível de indivíduos (GONÇALVES; UHEYAMA, 2012; ALVES, 2016). Devido as características de identificação de aspectos comportamentais, técnicas de Aprendizado de Máquina (*Machine Learning* (ML)), têm sido amplamente investigadas, tanto para a classificação de aspectos físicos (AUGUSTO *et al.*, 2007; OMATU *et al.*, 2009; CARVALHO *et al.*, 2010) quanto para aspectos emocionais (KHANCHANDANI; HUSSAIN, 2009; ZHOU *et al.*, 2011; PETER; URBAN, 2012; MANO *et al.*, 2016a; Mano, 2018), sendo tratadas como um problema de classificação comumente supervisionado. Contudo, recentemente, estudos têm explorado a abordagem multimodal, isto é, fusão de informações de diferentes modalidades (MANO *et al.*, 2016b; PORIA *et al.*, 2016; NETO *et al.*, 2017; PORIA *et al.*, 2017; GONÇALVES *et al.*, 2017a), relatando bons resultados dessa técnica aplicada ao desafio de classificação de aspectos comportamentais.

Em vista disso, uma solução que se apresenta é desenvolver sistemas que consideram as mudanças relativas ao comportamento, tanto físico quanto emocional, oferecendo a cada usuário um sistema que se ajuste de acordo com suas preferências, necessidades e intenções de uso (NERIS; BARANAUSKAS, 2010; ALVES, 2016). Nesse sentido, temos a hipótese de que a utilização da abordagem multimodal, isto é, baseada em múltiplos sensores (p. ex. câmera, acelerômetro, movimento, sensores residenciais, entre outros), permite a análise do estado físico e emocional de indivíduos de maneira não intrusiva, modularmente flexível e com maior precisão. Ainda, acreditamos que tal abordagem possa ser incorporada no ambiente residencial, por meio de uma arquitetura HSH, para o monitoramento de aspectos físicos e emocionais do usuário, visto sua estrutura multimodal.

1.1 Objetivos

A abordagem proposta investigou a identificação dos padrões do cotidiano, reconhecimento de emoções e estratégias de aprendizado inteligente das atividades e situações, de modo a fornecer um *feedback* do estado físico e emocional do usuário no ambiente residencial. Esta Tese de Doutorado teve como objetivo desenvolver e avaliar: **i.)** um modelo para identificar, de maneira inteligente e não intrusiva, o estado físico e emocional do cotidiano do usuário; **ii.)** a abordagem multimodal na classificação de aspectos físicos e emocionais do indivíduo, para isso foi realizado o estudo da face, do tom de voz e descrição da fala; **iii.)** uma arquitetura de comunicação e processamento para o ambiente HSH e; **iv.)** um sistema de monitoramento e dispositivo de baixo custo para saúde em casas inteligentes, de modo a facilitar a aquisição e interação de pessoas com deficiências, mobilidade reduzida e idosos com essa nova tecnologia.

1.2 Estrutura da Tese

O restante desta Tese está estruturado da seguinte maneira. No Capítulo 2, é apresentada a fundamentação teórica com o propósito de fornecer subsídios para o entendimento desta

pesquisa, abordando brevemente conceitos referente: i.) a IoT aplicada em ambientes HSH; ii.) ao conhecimento prévio sobre emoções, visto o caráter subjetivo para sua identificação, e; iii.) a inteligência artificial para a classificação de aspectos comportamentais do usuário, com foco na abordagem multimodal. O Capítulo 3 apresenta uma síntese do levantamento bibliográfico realizado, que aborda trabalhos e problemas de classificação de aspectos físicos e emocionais, além de ambientes HSH. Nos Capítulos 4 e 5, são apresentadas as soluções para classificação de aspectos emocionais e físicos do usuário, respectivamente. O Capítulo 6 apresenta a arquitetura para o monitoramento residencial para ambientes HSH. Por fim, no Capítulo 7, são apresentadas as conclusões desta pesquisa e trabalhos futuros, além das principais contribuições científicas desta Tese.

FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, é apresentada a fundamentação teórica para fornecer subsídios aos capítulos posteriores. Inicialmente, é apresentado um embasamento teórico referente a IoT e o ambiente HSH. Em seguida, é realizado um estudo para conhecimento prévio sobre emoções e a Teoria Componencial das Emoções. Posteriormente, são apresentados os conceitos de ML e a abordagem multimodal para a classificação do comportamento do usuário. Por fim, apresentadas as considerações finais deste capítulo.

2.1 Internet das Coisas aplicada na Saúde em Casas Inteligentes

A Internet das Coisas é um fenômeno tecnológico proveniente de empreendimentos inovadores e conceitos em tecnologia de informação e comunicação, associados com conectividade em Computação Ubíqua, Computação Pervasiva e Ambiente Inteligente (KÜHNER, 2007) como: identificação, localização, sensoriamento e conectividade:

- *Computação Ubíqua*: utiliza os conceitos de pervasividade aliados ao conceito de mobilidade, onde o sistema ubíquo pode estar em todo lugar de forma “invisível” ao usuário, porém, não restringe o seu campo de ação a um só ambiente, uma vez que pode estar em movimento constante juntamente com o usuário (mobilidade);
- *Computação Pervasiva*: implica que o dispositivo computacional está embarcado no ambiente de forma “invisível” para o usuário. Este conceito permite que sistemas computacionais tenham a capacidade de obter informações do ambiente ao qual estão inseridos e, dinamicamente, executar ações para atender dispositivos e usuários;
- *Inteligência Ambiental*: significa a capacidade dos objetos registrarem as mudanças no ambiente físico e, assim, interagir ativamente em um processo.

Normalmente, os objetos que cumprem estes requisitos são chamadas de “objetos inteligentes” (AARTS; GROTENHUIS, 2009). A IoT é definida como a capacidade de objetos inteligentes se comunicarem entre si e a capacidade de construção de redes de “coisas”.

Outra explicação para o conceito de IoT, é de que objetos podem apresentar características que permitem sua utilização para construir uma rede dinâmica conectada pela Internet (PEÑA-LÓPEZ *et al.*, 2005), tornando a Internet um espaço virtual online para um sistema embutido no mundo físico real. Características típicas da IoT são: i.) criar novas redes independentes que operam com suas próprias infraestruturas; ii.) implementação de novos serviços e; iii.) possibilidade de aplicar novas e diferentes modalidades de comunicação entre pessoas-“coisas” e “coisas”-“coisas” (DOHR *et al.*, 2010; GUBBI *et al.*, 2013). Além disso, objetos inteligentes podem estar ativos (tomada de decisão local) ou passivos (dados do sensor são armazenados e podem ser utilizados para leitura). Os principais objetivos na utilização da IoT é fornecer informações relevantes, quando e onde forem necessários, no sentido de preencher a lacuna entre a Web e o mundo real (GUBBI *et al.*, 2013).

Nesse sentido, uma proposta de aplicação dos conceitos de IoT surge para Saúde em Casas Inteligentes, onde sistemas computacionais são utilizados para apoiar as pessoas idosas e com necessidades especiais em sua rotina diária. O principal objetivo é manter e promover a independência das pessoas, aumentando a segurança no seu estilo de vida e no seu ambiente residencial (MANO *et al.*, 2016a). A necessidade de tais aplicações surge da evolução demográfica nos países industrializados, onde a expectativa de vida da população aumenta e a taxa de natalidade diminui (MANO *et al.*, 2016b). Tais circunstâncias exigem soluções inovadoras e de baixo custo para manter os gastos com a saúde dentro dos limites da possibilidade econômica da sociedade (DOHR *et al.*, 2010).

Aplicações HSH incluem serviços, produtos e conceitos para aumentar a qualidade de vida, bem-estar e segurança das pessoas. O principal objetivo é fornecer benefícios para o indivíduo (aumentando a segurança e bem-estar), a economia (maior eficácia dos recursos limitados) e a sociedade (melhores condições de vida) (TAKÁCS; HANÁK, 2007; DOHR *et al.*, 2010). Os campos de necessidades de pessoas com alguma limitação e idosas em todas as aplicações são: i.) Saúde; ii.) Segurança; iii.) Paz de espírito; iv.) Independência; v.) Mobilidade e; vi.) Interação Social (MANO *et al.*, 2016b).

Como pode-se notar, o escopo de aplicações nestes domínios é muito amplo. Por esta razão, ambientes HSH são estruturados em três níveis distintos: Hardware (sensoriamento e redes sem fio), *Middleware* (captura de dados, segurança de dados e integração) e Serviços (processamento de sinais biológicos, processos centrados em aplicações e serviços) (KUNZE *et al.*, 2008). HSH tem uma forte relação com a “Inteligência Ambiental”, sendo uma tecnologia de ponta para a IoT. Por exemplo, para ativar o aumento da segurança e bem-estar em sua casa, a casa tende a se tornar inteligente com a ajuda da “rede das coisas”.

O cenário HSH caracteriza-se por estar conectado, sensível ao contexto pessoal, adaptável

e de antecipação. O que é suposto, ser capaz de fornecer todas as características necessárias para um ambiente assistido. No que diz respeito aos campos de necessidades para as pessoas com alguma deficiência, mobilidade reduzida e idosos, é possível abranger diversos campos por meio da IoT. O monitoramento de doenças crônicas (saúde), disposição com alimentos frescos e sistemas de alarme (segurança), serviços de lembrete (paz de espírito) e permitir a comunicação pessoa-a-pessoa, por exemplo, com os familiares e/ou amigos (interação social). Ainda, a comissão da União Europeia incentiva essa hipótese, dizendo: “O escopo de aplicações da Internet das Coisas é esperado para contribuir grandemente a abordar desafios sociais de hoje” (II, 2014).

2.2 Conhecimento prévio sobre emoções - Teoria Componencial das Emoções

A emoção é uma reação complexa que envolve todo o organismo do indivíduo, tendo uma relação direta com as suas necessidades, metas, valores e bem-estar em geral. No estudo geral das emoções, diferentes componentes são considerados em pesquisas, e trazem a tona a discussão sobre o que as emoções são e o que precisa ser estudado quando se quer determinar seu significado (FONTAINE *et al.*, 2002; SCHERER, 2005; ALVES, 2016). Alguns estudos se concentram nos antecedentes das emoções e a avaliação de situações (FRIJDA, 1986; PARKINSON, 1995; SCHERER, 2001) e em termos de bem-estar de um indivíduo (ELLSWORTH; SCHERER, 2003). Outros pesquisadores abordam os padrões da reação psicofisiológica (EKMAN; FRIESEN, 1981; STEMMLER, 2003).

Diversos modelos emocionais foram propostos ao longo da história dos estudos sobre emoções. No entanto, dentre estes modelos, existem duas vertentes principais para a modelagem emocional: teorias dimensionais das emoções e teorias cognitivas das emoções (RUSSELL, 1980). A primeira, resultou nos modelos contínuos, ao passo que a segunda nos modelos discretos ou categóricos. Enquanto os modelos discretos preocupam-se em agrupar rótulos semelhantes sob uma mesma categoria emocional, os modelos contínuos visam descrever a relação entre tais categorias (ALVES, 2016).

Dentre os modelos contínuos propostos, o mais atual e difundido é o Modelo Circumplexo proposto por Scherer (2005) (Figura 1) que argumenta que todas as emoções se encontram em um espaço bidimensional contínuo, dividido em 4 eixos principais, em que as dimensões são (ALVES, 2016):

- *Valência*: corresponde ao tipo de emoção e representa como um ser humano se sente (eixo X - positiva ou negativa);
- *Excitação*: corresponde à intensidade da emoção e mede a propensão dos seres humanos em realizar uma ação desencadeada pelo estado emocional (eixo Y - ativo ou passivo,

língua ou cultura das pessoas envolvidas na comunicação. A principal característica do modelo proposto por Ekman (1973) e Ekman (2006) é a semelhança de como as pessoas utilizam tal esquema para descrever demonstrações emocionais observadas no cotidiano. O modelo permite facilmente associar as emoções com expressões faciais que as representam, pois implicam em alterações nas expressões que acompanham a experiência emocional do usuário (MANO *et al.*, 2019a). Por exemplo, a face sofre mudanças de acordo com o grau de excitação; em termos de respostas emocionais têm-se, por exemplo, um olhar de ódio, o franzir a testa, comprimir os lábios ou até mesmo um sorriso. Todas as outras categorias emocionais são, então, construídas a partir de combinações dessas emoções básicas.

As emoções analisadas nesse estudo são baseadas no modelo proposto por Ekman (1973) e Ekman (2006) com adição da expressão Neutro (como marco zero para as emoções), por se tratar de uma abordagem mais recente (MANO *et al.*, 2019a; MANO *et al.*, 2019b). Tais expressões são apresentadas na Figura 2.

Figura 2 – Emoções básicas propostas por Ekman (1973) e Ekman (2006), com a expressão Neutro.

Fonte: Ekman (1973), Ekman (2006), Mano *et al.* (2019a), Mano *et al.* (2019b).

Diante do exposto, a Teoria Componencial das Emoções (STEMMLER, 2003; SCHERER, 2005; NIEDENTHAL; KRAUTH-GRUBER; RIC, 2006; ALVES, 2016) entende as emoções como sendo níveis variáveis de mudanças inter-relacionadas entre um conjunto de cinco componentes: avaliações cognitivas, experiências subjetivas, tendências comportamentais, reações fisiológicas e expressões motoras. Mahlke e Minge (2008) destacam que uma abordagem baseada em componentes oferece uma maneira abrangente para entender os diferentes aspectos das emoções em todo o tipo de situações. Assim, diferentes aspectos podem ser utilizados para a análise emocional sem, no entanto, deixar de considerar a relação entre os diferentes componentes (SCHERER, 2001). A seguir, são detalhados os cinco componentes da Teoria Componencial das Emoções (ALVES, 2016):

1. *Avaliação Cognitiva*: está ligada à interpretação de uma situação e contribui para o desenvolvimento das emoções. O indivíduo avalia constantemente o mundo ao seu redor e procura perceber as qualidades afetivas de eventos, objetos e recursos (FONTAINE *et al.*, 2002). Essa constante avaliação implica em utilizar recursos sensoriais, perceptivos e cognitivos para interpretação emocional do usuário (SCHERER, 2005; MAHLKE; MINGE, 2008).

2. *Experiência Subjetiva*: é responsável pela avaliação de um episódio emocional, ou seja, da experiência subjetiva consciente que se relaciona com a capacidade do indivíduo de regular as emoções (DESMET, 2003). De acordo com Scherer (2005), não há nenhum método objetivo capaz de avaliar os sentimentos subjetivos, pois a única maneira de se obter tais informações é perguntando ao próprio indivíduo.
3. *Tendências Comportamentais*: preparam e orientam o organismo para, ou durante, uma determinada ação (FONTAINE *et al.*, 2002). Esse componente aponta a direção e a energia de repostas emocionais específicas para o desempenho de tarefas que preparam o indivíduo para uma determinada situação (EKMAN, 2006).
4. *Reações Fisiológicas*: são aspectos fisiológicos que permite ao usuário revelar espontaneamente e inconscientemente suas emoções (SCHERER, 2005; ALVES, 2016). Esse componente é responsável por regular e controlar as funções do corpo, tais como: respiração, digestão, circulação, entre outros.
5. *Expressões Motoras*: são responsáveis por comunicar aspectos comportamentais (SCHERER, 2005; MAHLKE; MINGE, 2008; XAVIER; GARCIA; NERIS, 2012). Esse componente implica em alterações nas expressões faciais, vocais e gestuais que acompanham a experiência emocional do usuário.

É importante notar que ambas as representações de emoções, contínuas e categóricas, têm sido utilizadas em várias aplicações computacionais com bom desempenho (SCHERER, 2005; MANO *et al.*, 2015; ALVES, 2016; MANO *et al.*, 2016a; MANO *et al.*, 2019a; MANO *et al.*, 2019b).

2.3 Inteligência Artificial para a classificação do estado físico e emocional

O constante aumento da capacidade de processamento computacional e os estudos na área de Interação Humano-Computador (IHC) vêm motivando pesquisadores a tentar identificar aspectos individuais de usuários, em especial relacionados a questão física e emocional, e utilizar essa informação como base para a tomada de decisão, análise de satisfação e comportamento na execução de tarefas (FILHO *et al.*, 2013). De fato, procedimentos de classificação têm ajudado na análise de repostas físicas e emocionais, auxiliando no diagnóstico da depressão, acidentes residenciais, mudança de comportamento, entre outros aspectos, fornecendo a oportunidade para análise em diversos contextos (ALVES, 2016).

Pesquisas na área de ML têm gerado métodos que “aprendem” a partir de exemplos, e que podem ser utilizados para extrair padrões para a classificação (DUDA; RICHARD, 1973; CLARK; NIBLETT, 1989; CRISTIANINI; SHAW-TAYLOR, 2000; UEYAMA *et al.*, 2014;

ALVES, 2016). Tais técnicas de classificação estão cada vez mais sendo aplicadas tanto para análise de aspectos físicos do usuário (OMATU *et al.*, 2009; MAIA *et al.*, 2015; MANO *et al.*, 2016b; NETO *et al.*, 2019; OLIVEIRA *et al.*, 2018) quanto para a análise de respostas emocionais (ØHRN; ROWLAND, 2000; CHANEL *et al.*, 2009; ZHOU *et al.*, 2011; GONÇALVES *et al.*, 2013a; RAMAKRISHNAN; EMARY, 2013; ALVES, 2016). Abaixo segue uma súmula conceitual de alguns algoritmos de classificação utilizados no reconhecimento de padrões dos usuários (ALVES, 2016) e exploradas nesta Tese:

- *Decision or Regression Tree* – A Árvore de Decisão (Árv. Dec.) ou regressão, aplicada nos trabalhos de Bailenson *et al.* (2008) e Peter e Urban (2012), utiliza uma árvore para realizar a classificação ou estimar de um valor de um dado teste. Um dos algoritmos mais famosos para gerar essas árvores é o C4.5, cuja árvore cresce a partir de uma abordagem “dividir e conquistar” e é ramificada utilizando o atributo que obtiver o melhor ganho de informação (DUDA; RICHARD, 1973).
- *Fuzzy Logic* - A Lógica Fuzzy, utilizada nos trabalhos de Katsis *et al.* (2008) e Peter e Urban (2012), é capaz de processar dados incompletos e de prover soluções aproximadas para problemas que outros métodos encontram dificuldades em solucionar (CLARK; NIBLETT, 1989; KATSIKIS *et al.*, 2008).
- *k-Nearest Neighbor classification (kNN)* - Classificação dos *k*-vizinhos mais próximos. O kNN, utilizado em alguns estudos (BAILENSEN *et al.*, 2008; YUAN *et al.*, 2010) para o reconhecimento emocional, é uma abordagem de classificação não-paramétrica, isto é, não assume nenhuma distribuição de probabilidade para os dados *a priori*. O algoritmo kNN encontra um grupo de *k* objetos no conjunto de treino que estão mais próximos do objeto de teste e o classifica com a classe predominante dos seus *k*-vizinhos (FACELI, 2011).
- *Bayesian Networks* – A Rede Bayesiana é um modelo de grafo probabilístico, aplicado nos trabalhos de Chanel *et al.* (2009), Peter e Urban (2012) e Ramakrishnan e Emary (2013), e que representa um conjunto de variáveis aleatórias e suas dependências condicionais como um grafo acíclico¹ direto (FACELI, 2011). Por exemplo, o algoritmo Naive Bayes (N. Bayes) pode representar a relação entre os sintomas e as doenças, isto é, dado os sintomas, pode-se inferir a probabilidade de várias doenças.
- *Support Vector Machine (SVM)* - O algoritmo SVM, utilizados por Littlewort *et al.* (2011), Peter e Urban (2012) e Ramakrishnan e Emary (2013) para o reconhecimento de emoções, é um dos métodos mais robustos e acurados entre os algoritmos conhecidos. Segundo Faceli (2011), o objetivo do SVM é encontrar a melhor função para distinguir duas ou mais classes dos dados de treinamento. A garantia da melhor função é atingida, pois o

¹ Acíclico - aquilo que não é cíclico, ou seja, sua ocorrência é aleatória, mas pode se prever a sua frequência, pelo uso da estatística.

algoritmo maximiza a margem entre as classes e oferece melhor habilidade de generalização (CLARK; NIBLETT, 1989; LITTLEWORT *et al.*, 2011; RAMAKRISHNAN; EMARY, 2013).

No entanto, pesquisas recentes (D'MELLO; KORY, 2015; PORIA *et al.*, 2017; GONÇALVES *et al.*, 2017a; FIORI; VESELY-MAILLEFER, 2018) apresentam o conceito de classificação multimodal, isto é, a combinação de respostas de sensores e/ou algoritmos para o reconhecimento de padrões, em particular no reconhecimento de emoções.

A classificação multimodal pode ser visto como a fusão de informações de diferentes modalidades, isto é, o processo de combinar dados coletados de várias modalidades para tarefas de análise e classificação (PORIA *et al.*, 2017), fornecendo informações adicionais com o aumento da precisão do resultado geral (D'MELLO; KORY, 2015). No entanto, como as informações são adquiridas de modalidades diferentes, é necessário considerar o período para a fusão multimodal em diferentes níveis. Segundo Poria *et al.* (2017), os níveis de fusão mais difundidos são: em nível de característica, em nível de decisão, fusão híbrida e fusão em nível de modelo. Tais níveis de fusão são descritos a seguir:

- *Nível de característica*: realiza a fusão dos recursos extraídos de várias modalidades, como por exemplo recursos visuais, sonoros e textuais, como um vetor geral para a análise. A vantagem da fusão em nível de característica é que a correlação no estágio inicial pode, potencialmente, facilitar a tarefa de classificação. A desvantagem desse processo de fusão é a sincronização de tempo, já que os recursos obtidos pertencem a diversas modalidades e podem discordar em muitos aspectos (ROSAS; MIHALCEA; MORENCY, 2013; SARKAR *et al.*, 2014; PORIA *et al.*, 2015).
- *Nível de decisão*: as características de cada modalidade são tratadas e classificadas, de forma independente, e os resultados são fundidos como um vetor de decisão para obter a decisão final. A vantagem do nível de decisão é que cada modalidade pode utilizar seu melhor classificador ou modelo para aprender suas características. No entanto, como desvantagem, o processo de aprendizagem de todos esses classificadores no estágio de fusão no nível de decisão torna-se custoso computacionalmente, visto o processo de aprendizagem de todos os classificadores no estágio de fusão no nível de decisão (ALAM; RICCARDI, 2014; CAI; XIA, 2015).
- *Fusão híbrida*: é a combinação dos níveis anteriores (nível de característica e nível de decisão). Este foi proposto na tentativa de explorar as vantagens das estratégias e superar as desvantagens de cada nível (WÖLLMER *et al.*, 2013b; PORIA; CAMBRIA; GELBUKH, 2015; PORIA *et al.*, 2016).

- *Nível de modelo*: é uma técnica que usa a correlação entre dados observados, em diferentes modalidades, com uma fusão descontraída dos dados, baseada na natureza básica dos métodos e no espaço do problema (METALLINOU *et al.*, 2012; WÖLLMER *et al.*, 2013a).

Considerando os aspectos multidimensionais do espaço semântico de Scherer (2005) e a Teoria Componencial das Emoções (SCHERER, 2001), o uso da abordagem multimodal faz sentido por considerar a abrangência que o estudo reflete sobre os aspectos físicos e emocionais do indivíduo (SCHERER, 2005; MAHLKE; MINGE, 2008; XAVIER; GARCIA; NERIS, 2012). A Figura 16 ilustra a classificação multimodal no nível de característica, objeto de estudo desta Tese, para classificação das emoções do usuário considerando múltiplos sensores e Inteligência Artificial.

Figura 3 – Cenário de classificação multimodal da emoção considerando abordagem multimodal no nível de característica.

Fonte: Poria *et al.* (2015).

Vale salientar, que a abordagem multimodal adotada, não fica limitada ou estática a sensores específicos, podendo ser adicionados outros sensores a estrutura do modelo (linha tracejada da Figura 16), tornando-a dinâmica aos sensores utilizados no monitoramento dos usuários.

2.4 Considerações finais

Neste capítulo, foram abordados os principais conceitos relacionados a esta Tese. Primeiramente, apresentamos brevemente os conceitos de IoT e sua utilização para prover ambientes

HSH. Em seguida, visto a subjetividade relacionada às emoções, apresentamos conceitos prévios sobre emoções e a Teoria Componencial das Emoções. Por fim, apresentamos, brevemente, os conceitos das abordagens utilizadas para a classificação de aspectos físicos e emocionais do usuário, com ênfase na abordagem multimodal.

TRABALHOS RELACIONADOS

Este capítulo apresenta uma investigação do estado da arte referente as metodologias de classificação de aspectos físicos e emocionais do indivíduo, além de apresentar a tendência de crescimento dos sistemas de monitoramento residencial, principalmente voltados a saúde do usuário. Inicialmente, abordamos trabalhos e problemas para a classificação emocional, com foco nas características da face, voz e descrição da fala, além do uso multimodal para a classificação emocional. Em seguida, são apresentados trabalhos referentes a classificação de aspectos físicos, referentes a classificação de atividades e qualidade do ambiente para o repouso. Posteriormente, discutimos estudos referentes ao ambiente HSH. Por fim, são apresentadas as considerações finais deste capítulo.

3.1 Classificação de aspectos emocionais

Pesquisas (GONÇALVES *et al.*, 2011; ALVES, 2016; PORIA *et al.*, 2017; GONÇALVES *et al.*, 2017a) na área de IHC, inteligência artificial e processamento de sinais têm gerado métodos capazes de extrair padrões de aspectos emocionais, sendo utilizados em diversos domínios e aplicações, como a tomada de decisão sobre questões de saúde e bem-estar dos usuários. No entanto, a classificação da emoção é uma tarefa complexa, visto o viés subjetivo que norteia a emoção humana e seu significado.

No primeiro contato de interação social, as expressões faciais são as pistas primárias e o principal canal para formar uma impressão do atual estado emocional do sujeito. As pessoas de diferentes etnias, regiões e culturas compartilham diferentes representações faciais para informar as emoções, voluntariamente ou não. Ekman (1973), Ekman e Keltner (1997), considerados os pioneiros no estudo das emoções, argumentam ser possível detectar as emoções básicas a partir de pistas de expressões faciais. Nesse sentido, Friesen e Ekman (1978) desenvolveu um sistema de codificação facial, denominado *Facial Action Coding System* (FACS). O FACS é amplamente utilizado, pois se baseia na reconstrução de expressões faciais baseadas no movimento do

músculo facial. Os códigos FACS são usados para inferir emoções usando uma variedade de recursos disponíveis, como o *FACS Investigators' Guide* (EKMAN; FRIESEN; HAGER, 2002), o banco de dados interpretativo FACS (EKMAN; ROSENBERG; HAGER, 1998) e um grande corpo de pesquisa empírica (EKMAN, 1997).

De fato, pesquisas apontam para o bom desempenho do reconhecimento facial para a classificação da emoção. Mano *et al.* (2015), Alves (2016), Mano (2018) utilizam características geométricas para análise da expressão facial. Os métodos baseados em características geométricas permitem modelar a forma e a localização dos componentes faciais (incluindo a boca, olhos, sobrancelhas e nariz), utilizando pontos e elementos geométricos, como ângulos, distâncias e áreas para representar a geometria da face. Os estudos utilizam ML e o conceito de conjunto de classificadores na classificação da emoção, na tentativa de suavizar o erro que um único classificador pode cometer.

Outros métodos de reconhecimento facial têm apresentado resultados promissores na literatura, é o caso da abordagem baseada em Redes Neurais Convolucionais (*Convolutional Neural Network* (CNN)) e arquiteturas para reconhecimento emocional baseado na face (PONS; MASIP, 2018; JAIN *et al.*, 2018; O'TOOLE *et al.*, 2018). As CNNs são baseadas no aprendizado de transferência de informação de dados, motivados pela necessidade de processamento de dados cada vez maiores no campo de análise emocional. A principal vantagem desta abordagem é o fato da não exigência de conhecimento prévio sobre emoções, a capacidade de tratamento de dados “ruídosos” e a estratégia progressiva para aperfeiçoar redes de aprendizagem profunda. De outra perspectiva, Cruz-Albarran *et al.* (2017) propõe uma metodologia baseada em informações térmicas para delimitar a região de interesse das expressões faciais, e estimar as emoções dos indivíduos através da tendência térmica que acompanha determinada emoção.

Apesar da ampla utilização da face para o reconhecimento de emoções, outros aspectos também são explorados. O trabalho proposto por Khanchandani e Hussain (2009) utiliza *features* prosódicas: frequências formantes, entropia, variância, mínimos, mediana e corretor de predição linear para a classificação de sinais de fala com carga emocional. O estudo utiliza gravações de áudio e determinam um nível mínimo de energia no sinal da fala para análise e classificação, por meio de redes neurais *Multilayer Perceptron* (MLP). Similarmente, Neto *et al.* (2018) utilizam coeficientes cepstrais e energia para processamento do sinal da fala, e por meio de técnicas de ML, realizam a classificação emocional da fala do usuário.

De outra perspectiva, o trabalho proposto por Wang *et al.* (2015) realiza o estudo de características visuais e textuais usadas na análise de sentimento em redes sociais. Para a construção do modelo foi utilizada uma técnica ML não supervisionado. Enquanto esta técnica possibilita a construção de grandes bases de dados por não exigir a rotulação manual de indivíduos, as *hashtags* foram consideradas como maneira de anotar as postagens e, sendo assim, tais conteúdos foram considerados para a predição. Por outro lado, LiKamWa *et al.* (2013) realizaram um estudo que se diferencia de outras aplicações de detecção de humor, pois não

utilizam dados de sensores físicos e sim de sensores lógicos. O estudo para o reconhecimento de humor é baseado em seis tipos de informações: SMS, e-mail, registro de ligações, uso de aplicações, navegação na Web e localização, permitindo a detecção automática de humor para a criação de novos aplicativos e a melhoria de outros existentes.

Modelos baseados em reações fisiológicas também têm ganhado destaque na literatura. [Peter e Urban \(2012\)](#) apresentaram uma abordagem geral de reconhecimento de emoções que utiliza sensores fisiológicos. O sistema coleta dados de alguns tipos de sensores: atividade eletrodérmica, temperatura do ar atmosférico e variabilidade da frequência cardíaca, para o reconhecimento emocional em sistemas de interação e ambiente educacional. Similarmente, [Zhou et al. \(2011\)](#) desenvolveram um estudo com o objetivo de prever diferentes estados emocionais do usuário por meio de reações fisiológicas. O estudo foi desenvolvido para provocar respostas emocionais com fotos de emoções padronizadas, ao passo que algumas reações fisiológicas foram medidas. A pesquisa desenvolvida utilizou apenas sensores que analisavam as reações fisiológicas e utilizam a fala do usuário para a definição das classes (saídas), por exemplo, quando o indivíduo dizia: “Ao olhar essa imagem, eu fico triste”.

Recentemente, estudos têm aplicado a abordagem multimodal para a classificação de sentimentos e emoções, sendo um pré-requisito importante para a implementação bem-sucedida da interação entre o agente e o usuário ([PORIA et al., 2017](#)), fornecendo uma avaliação mais precisa dos estados emocionais do que quando se utiliza a abordagem unimodal. [Poria et al. \(2016\)](#) propõem uma metodologia para análise de sentimentos multimodais, que consiste em extrair sentimentos de vídeos da Web, utilizando as modalidades de áudio, vídeo e texto como fontes de informação. [Gonçalves et al. \(2017a\)](#) e [Gonçalves et al. \(2017b\)](#) propõem um estudo que leva em conta múltiplos sensores para que se possa determinar o estado emocional do usuário no momento da interação. A abordagem multissensorial proposta inclui várias entradas de usuários (como fala, movimentos faciais e atividades cotidianas) e utiliza ML para mapear essas diferentes respostas em um ou mais estados emocionais.

O Quadro 1 apresenta uma visão geral das pesquisas encontradas na literatura e detalhadas acima. De acordo com os objetivos desta Tese, a análise considera as seguintes características: i.) ambiente de HSH proposto para monitoramento da saúde; ii.) componente de análise emocional; iii.) abordagem de classificação emocional; iv.) análise de aspectos físicos para cuidados de saúde, e v.) interação não intrusiva com o usuário.

Nota-se a grande gama de trabalhos que tratam do reconhecimento emocional. No entanto, é possível perceber que não há consenso referente ao componente emocional e a abordagem de classificação para determinar aspectos emocionais. Tal característica deve-se ao caráter subjetivo e aos diversos aspectos relacionados às emoções. Todavia, segundo o modelo proposto por [Scherer \(2005\)](#), durante o processo de avaliação emocional, deve-se considerar a associação das emoções aos quatro domínios da experiência emocional (Valência, Excitação, Sentimento de Controle e Facilidade de Conclusão do Objetivo), juntamente com as mudanças relacionadas

Quadro 1 – Análise dos trabalhos identificados na literatura referente a classificação de aspectos emocionais.

Trabalhos relacionados	Ambiente HSH	Componente emocional	Abordagem de classificação	Aspectos físicos	Interação não intrusiva
Ekman (1973)	X	face	-	X	-
Friesen e Ekman (1978)	X	face	-	X	-
Ekman e Keltner (1997)	X	face	-	X	-
Ekman (1997)	X	face	-	X	-
Ekman, Rosenberg e Hager (1998)	X	face	-	X	-
Ekman, Friesen e Hager (2002)	X	face	-	X	-
Khanchandani e Hussain (2009)	X	voz	MLP	X	✓
Zhou <i>et al.</i> (2011)	X	sinais fisiológicos	kNN e Árv. Dec.	X	X
Peter e Urban (2012)	X	sinais fisiológicos	Diversos	X	X
LiKamWa <i>et al.</i> (2013)	X	sensores lógicos	Estatístico	X	X
Mano <i>et al.</i> (2015)	X	face	Comitê	X	✓
Wang <i>et al.</i> (2015)	X	texto e imagem	SVM	X	✓
Alves (2016)	X	face	Comitê	X	✓
Poria <i>et al.</i> (2016)	X	face, voz e contexto	Multimodal	X	✓
Cruz-Albarran <i>et al.</i> (2017)	X	face	Hierárquico	X	✓
Gonçalves <i>et al.</i> (2017a)	X	face, fala, atividade	Multimodal	X	X
Gonçalves <i>et al.</i> (2017b)	X	face, fala, atividade	Multimodal	X	X
Pons e Masip (2018)	X	face	CNN	X	✓
Jain <i>et al.</i> (2018)	X	face	CNN	X	✓
O'Toole <i>et al.</i> (2018)	X	face	CNN	X	✓
Neto <i>et al.</i> (2018)	X	voz	kNN	X	✓
Proposta	✓	face, voz e contexto	Multimodal	✓	✓

aos componentes da Teoria Componential das Emoções (ver Seção 2.2). Em virtude disso, a abordagem multimodal (PORIA *et al.*, 2017; PORIA *et al.*, 2016; GONÇALVES *et al.*, 2017a; GONÇALVES *et al.*, 2017b) contribui com uma maneira para entender a análise emocional nos diversos tipos de situações.

Entretanto, considerando os objetivos desta Tese, observa-se a escassez de estudos do reconhecimento de emoções para ambientes HSH, visto que aspectos emocionais têm um grande impacto no que diz respeito à saúde e bem-estar dos indivíduos, principalmente a população idosa (MANO *et al.*, 2016a). Tal característica é satisfeita neste trabalho, por meio de um sistema para o monitoramento de emoções no ambiente residencial, juntamente com a análise de aspectos físicos, fornecendo um sistema HSH para o cuidado residencial de pessoas com alguma deficiência, mobilidade reduzida e idosos.

3.2 Classificação de aspectos físicos

O monitoramento de aspectos físicos e/ou do ambiente é um instrumento indispensável quando se quer reduzir problemas de mobilidade e dependência (MANO *et al.*, 2016b). Estudos recentes (VALLABH; MALEKIAN, 2018; LAPIERRE *et al.*, 2018; CADDICK *et al.*, 2018) têm abordado questões relacionados à queda de idosos ou ambientes para o repouso, objetos desse estudo, visto que tais aspectos têm grande influência no cotidiano, na saúde e bem-estar dos indivíduos, principalmente pessoas com necessidades e idosos.

Augusto *et al.* (2007) apresentam uma solução para reduzir acidentes que idosos podem sofrer durante o período noturno, com foco na análise do ambiente, proporcionando um sistema não intrusivo. O sistema baseia-se na visão noturna para a análise do ambiente e, por meio de algoritmos baseados na noção de causalidade e espaço temporal, fornece informações e alertas sobre acidentes e/ou possíveis situações de risco do idoso. Similarmente, Omatu *et al.* (2009) apresentam uma arquitetura de rede e aproveitam os benefícios oferecidos por um ambiente inteligente, utilizando dados de diferentes sensores. O objetivo do sistema é garantir a segurança e fornecer serviços aos usuários através de técnicas de monitoramento com base no processamento de imagem. Neste caso, por meio de um sensor de imagem e algoritmos baseados na combinação de filtros e comparações de taxa de quadro, atuam de forma inteligente no ambiente, permitindo a detecção de movimentos e padrões nas atividades do usuário, como por exemplo a velocidade e a direção que ele se movimenta. Apesar do caráter não intrusivo, ambas as abordagens são baseadas no tratamento de imagens, o que pode ser custoso computacionalmente para o sistema e limita sua aplicação para ambientes totalmente pré-definidos.

Corchado *et al.* (2010) propõem um sistema de monitoramento para melhorar a qualidade de vida dos usuários que dependem de assistência médica, focados em sensores fisiológicos. O sistema implementa uma plataforma baseada em arquitetura orientada a serviços, e permite que sensores comuniquem-se de maneira distribuída, independentemente das restrições de tempo e localização. Na mesma linha de pesquisa, Maia *et al.* (2015) desenvolveram uma plataforma de *middleware* para IoT baseada na Web, e conecta médicos e pacientes utilizando sensores corporais e atuadores. O objetivo principal do sistema é o monitoramento de pessoas em observação, uma vez que sugere sensores de respiração, um sensor de temperatura e um sensor de eletrocardiograma. Dessa forma, o médico pode avaliar o paciente em tempo real e remotamente. No entanto, considerando o cenário de uma HSH, o monitoramento por tais sensores é considerado extremamente intrusivo e desconsidera a qualidade de vida e independência do usuário em seu cotidiano. Além disso, os autores não consideraram a análise das atividades ou mesmo do ambiente para o cuidado no ambiente residencial.

Neto *et al.* (2017) propõem e avaliam um sistema flexível que combina o paradigma *fog computing* e *publish/subscribe* para monitorar e gerenciar de maneira individualizada indivíduos com mobilidade reduzida. O sistema avalia o reconhecimento de indivíduos e a detecção de quedas, permitindo conectar novos dispositivos e aplicações de forma escalável a sua infraestrutura em tempo real, bem como um melhor aproveitamento dos recursos dos dispositivos por meio do *fog*. Similarmente, Mano *et al.* (2016b) discutem a detecção de quedas no ambiente HSH e propõem um sistema, baseado em Inteligência Artificial e sensores de dispositivo embarcado, para o monitoramento residencial de pessoas idosas ou com alguma mobilidade reduzida. No entanto, apesar de tratarem do monitoramento de atividades e da interação não intrusiva de sistemas, ambas as abordagens não consideram outros aspectos que podem influenciar a saúde ou cotidiano do usuário, como por exemplo questões relacionadas com aspectos emocionais e/ou ambiente em geral.

Chung e Oh (2006) apresentam um exemplo de monitoramento e acompanhamento do ambiente. O estudo apresenta um módulo para monitorar a qualidade do ar do ambiente domiciliar e faz o uso de sensores de umidade, temperatura, CO₂, poeira, entre outros sensores. O sistema é adicionado ao ambiente doméstico, que incluem câmera de vigilância e interruptores de alimentação para o controle de iluminação e consumo de energia, para melhorar a função de monitoramento do ambiente. No entanto, tal abordagem não abrange aspectos ligados ao indivíduo, mais precisamente para a qualidade do ambiente para o sono e atividades do usuário, tratando tão somente das questões do ambiente domiciliar.

No contexto de monitoramento de condições de indivíduos, Carvalho *et al.* (2010) propõem o Sistema de Assistência Domiciliar à Saúde, baseado em Computação Pervasiva, para a aquisição de dados de sensores fisiológicos e de movimento para a identificação da situação do paciente. Por outro lado, Mano *et al.* (2016a) propõem uma arquitetura IoT de monitoramento residencial individualizado para o acompanhamento do estado emocional do paciente, baseado em técnicas de ML. O sistema identifica o paciente e classifica seu estado emocional por meio da face, podendo emitir alertas referentes ao emocional do paciente em tratamento domiciliar. No entanto, ambas as abordagens para o monitoramento domiciliar, se tornam totalmente ou parcialmente intrusivas ao usuário, pois dependem da interação constante com os dispositivos, seja por meio dos sensores acoplados em seu corpo ou a interação para aquisição de imagens. Ainda, as abordagens não relacionam a qualidade do ambiente para o sono, o que pode influenciar no tratamento, na melhora e no cotidiano de indivíduos, principalmente a população idosa.

Por sua vez, especificamente referente ao monitoramento do repouso do indivíduo, Togeiro e Smith (2005) descrevem procedimentos, recomendações, achados e valores para diagnósticos utilizados em transtornos do sono. A polissonografia¹ e o teste múltiplo de latência do sono são considerados como métodos “padrão-ouro” na maioria dos transtornos do sono e narcolepsia. Questionários específicos incluindo avaliação da qualidade do sono, hiper-sonolência, transtornos respiratórios do sono e ritmo sono-vigília são utilizados, em geral, para mapear transtornos e estudos sobre o sono. Por outro lado, Loos, Ullrich e Kobayashi (2003) desenvolveram uma plataforma para unificar o *SleepSmart* (multi sensores na cama para detecção de posição do corpo, temperatura corporal, frequência cardíaca e respiratória) e o *Morpheus* (sistema que movimenta a cama para encorajar a pessoa a se movimentar e aliviar o ronco e a apneia) no intuito de um sistema mais uniforme no monitoramento e melhora da qualidade do sono. No entanto, apesar de tratarem questões relacionadas ao sono, ambas as abordagens necessitam da interação do usuário com os dispositivos, seja pela interação no preenchimento de questionários ou pela interação corporal, fornecendo ambientes totalmente intrusivos.

O Quadro 2 apresenta uma síntese das pesquisas encontradas na literatura e detalhadas anteriormente. De acordo com os objetivos desta Tese, as pesquisas foram analisadas conside-

¹ Polissonografia ou exame do sono é um teste multiparamétrico utilizado no estudo do sono e de suas variáveis fisiológicas.

rando as seguintes características: i.) ambiente de HSH proposto para monitoramento da saúde; ii.) processo de análise emocional para o cuidado; iii.) análise de aspectos físicos para cuidados de saúde, iv.) classificação da qualidade do ambiente para o sono, e; v.) interação não intrusiva com o usuário.

Quadro 2 – Análise dos trabalhos identificados na literatura referente a classificação de aspectos físicos.

Trabalhos relacionados	Ambiente HSH	Análise emocional	Análise de atividade	Qualidade do sono	Interação não intrusiva
Loos, Ullrich e Kobayashi (2003)	X	X	X	✓	X
Togeiro e Smith (2005)	X	X	X	✓	X
Chung e Oh (2006)	X	X	X	X	✓
Augusto <i>et al.</i> (2007)	✓	X	✓	X	✓
Omatu <i>et al.</i> (2009)	✓	X	✓	X	✓
Corchado <i>et al.</i> (2010)	✓	X	X	X	X
Carvalho <i>et al.</i> (2010)	✓	X	✓	X	X
Maia <i>et al.</i> (2015)	✓	X	X	X	X
Mano <i>et al.</i> (2016a)	✓	X	✓	X	✓
Mano <i>et al.</i> (2016a)	✓	✓	X	X	X
Neto <i>et al.</i> (2017)	✓	X	✓	X	✓
Proposta	✓	✓	✓	✓	✓

É possível observar um aumento significativo nos últimos anos sobre trabalhos que abordam o contexto de Saúde em Casas Inteligentes, tanto de monitoramento de atividades quanto de ambientes (MANO *et al.*, 2016b; MANO *et al.*, 2016a; OLIVEIRA *et al.*, 2018). Tais trabalhos contribuem com técnicas para o desenvolvimento de soluções capazes de identificar, por meio de sensores e/ou informações, um determinado aspecto do usuário, principalmente aspectos ligados as atividades e uma possível quedas de pessoas idosas.

Vale salientar, entretanto, a carência de trabalhos na área de Saúde em Casas Inteligentes que abordam conceitos relacionados a Computação Ubíqua, juntamente com questões referentes a qualidade do ambiente para o repouso dos usuários, visto que o ambiente é considerado um fator importante na análise do repouso, pois pode interferir diretamente na qualidade de vida e comportamento de usuários com idade elevada (CEOLIM; DIOGO; CINTRA, 2001; ANCOLI-ISRAEL, 2005). Ainda, é possível verificar a lacuna de trabalhos que abordam o uso de sensores e uma abordagem robusta para a classificação de ambientes para o repouso juntamente com aspectos físicos e emocionais do usuário. Tais características são satisfeitas neste trabalho por meio da coleta e análise de dados oriundos dos sensores de dispositivos móveis de maneira não intrusiva. Este estudo permite que sistemas computacionais utilizem a informação da qualidade do ambiente como característica complementar na análise do sono do usuário.

3.3 Arquitetura para Saúde em Casas Inteligentes

Estudos (CHAN *et al.*, 2008; ALAM; REAZ; ALI, 2012; TALAL *et al.*, 2019) referente a IoT e *eHealth/mHealth* geraram métodos para estabelecer padrões ambientais e/ou de usuário.

Em particular, no ambiente HSH, os avanços na área da saúde devem as aplicações do ambiente inteligente, bem como à inteligência artificial (Mano, 2018). Tais métodos são utilizados para monitorar e classificar fatores relacionados à saúde e ao bem-estar do indivíduo, onde sistemas computacionais são aplicados para apoiar as pessoas idosas e pessoas com necessidades especiais em sua rotina diária.

O Laboratório de Computação Mobile e Pervasiva da Universidade da Flórida, propõe o *Gator Tech Smart House* (HELAL *et al.*, 2005), um laboratório experimental e um ambiente real destinado a validar tecnologias e sistemas para casas inteligentes. O objetivo é criar ambientes do mundo físico com monitoramento remoto e serviços de intervenção, conduzindo pesquisas e desenvolvendo atividades projetadas para assistência aos idosos e pessoas com necessidades especiais. No mesmo sentido, o Instituto de Tecnologia da Georgia desenvolveu o *Aware Home Research Initiative* (ABOWD *et al.*, 2002). A pesquisa investiga como as novas tecnologias podem impactar a vida das pessoas no ambiente domiciliar em três áreas principais: i.) saúde e bem-estar; ii.) meios de comunicação e entretenimento digital, e; iii.) sustentabilidade. Ambos os trabalhos, permitiram a pesquisa em muitas áreas, tais como fornecimento de dados para residências, sensoriamento inovador, controle de infraestrutura e uma variedade de aplicações para auxiliar os moradores.

A respeito do acompanhamento das condições crônicas dos idosos, Mohamed *et al.* (2017) adotaram uma abordagem para HSH que se baseia em um sensor eletrocardiográfico, um sensor de eletromiografia, um sensor de temperatura e um sensor de movimento para monitorar sinais vitais de pacientes ou deficientes, em tempo real. Embora adote uma abordagem de IoT e HSH, o trabalho não aborda questões relacionadas com as emoções dos usuários, o que pode ter um efeito positivo ou negativo tanto na condição quanto no tratamento dos pacientes em um ambiente residencial.

O estudo realizado por Fernández-Caballero *et al.* (2014) delineia um sistema para melhorar a qualidade de vida e o cuidado dos idosos em residências, regulando suas emoções. O sistema utiliza câmeras e sensores corporais para monitorar as expressões faciais, gestuais e comportamentais dos idosos, bem como quaisquer dados fisiológicos necessários para o reconhecimento da emoção. Castillo *et al.* (2016) estabeleceu um sistema muito similar, com o acréscimo de um microfone para captar o que o usuário diz. No entanto, além de não tratarem nenhuma informação referente as condições físicas do usuário, as abordagens não realizam a análise de questões relacionadas ao processamento da informação e comunicação entre os dispositivos IoT, o que podem ter uma influência considerável na eficácia de cada sistema.

Mano *et al.* (2016a) estabeleceram uma arquitetura de monitoramento residencial de IoT individualizada para o monitoramento de estados emocionais, baseada em técnicas de ML para o reconhecimento da face do paciente por meio de camadas de processamento. De acordo com as taxas de requisição e processamento, o sistema adota camadas com maior poder de processamento para identificar os pacientes e avaliar seu estado emocional por meio de expressões faciais, sendo

capaz de emitir alertas e advertências em relação ao estado emocional do paciente em ambientes residenciais. No entanto, embora o trabalho aborde questões relacionadas à comunicação e à taxa de processamento, a tomada de decisão para as camadas de processamento é pré-definida, levando-se em conta apenas o processamento para a classificação das imagens, tornando o sistema estático a aplicação, e isso restringe sua flexibilidade.

Nota-se que os trabalhos são focados na infraestrutura IoT, não apenas ao software, mas considerando aspectos de redes e hardware. Augusto *et al.* (2018) apresentam uma metodologia centrada no usuário para desenvolver ambientes inteligentes. Os autores destacam a integração de tecnologias e propõem uma infraestrutura de software para fornecer serviços. O desafio abordado está relacionado à confiança em sistemas no mundo real, como a qualidade do desempenho, a correção comportamental e sua validação. Da mesma forma, Patel, Ali e Sheth (2017) descrevem uma abordagem inteligente para a análise de aplicativos de IoT e se concentram na automação das transições entre borda e nuvem em situações dinâmicas. Os autores descrevem uma arquitetura dividida em camadas, compreendendo diferentes componentes e atores que assumem papéis em soluções adaptáveis de análise de dados da IoT. No entanto, em ambos os trabalhos, qualquer avaliação é realizada para atestar como suas soluções impactam a infraestrutura de ambientes inteligentes.

Nesse mesmo sentido, Kang, Kim e Choo (2017) exploram uma arquitetura de três camadas para IoT, com base em diferentes cenários de recursos computacionais, para detecção e configuração dos dispositivos em tempo de execução. O foco está na descoberta dinâmica de dispositivos IoT e no gerenciamento de suas conexões. No entanto, além do estudo não abordar a flexibilidade do volume de dados para processamento da arquitetura, as configurações são limitadas ao hardware, o que pode tornar a solução dependente de dispositivos e/ou aplicativos específicos. Visto tal limitação, Neto *et al.* (2019) propõem uma arquitetura baseada em *micro-fog* e uma solução de descarregamento de carga de trabalho, visando garantir a provisão de recursos na arquitetura e sincronizar a complexidade do processamento de dados. No entanto, apesar de apresentar resultados relevantes quando implementada em dispositivos com recursos escassos para a identificação do indivíduo e detecção de quedas, o estudo não considera a abordagem emocional e/ou aspectos do ambiente residencial para a saúde dos usuários.

O Quadro 3 apresenta uma visão geral das pesquisas encontradas na literatura e detalhadas acima. De acordo com os objetivos desta Tese, a análise considera as seguintes características: i.) ambiente de HSH proposto para monitoramento da saúde; ii.) processo de análise emocional; iii.) análise de aspectos físicos para cuidados de saúde; iv.) comunicação entre camadas da rede e; v.) algoritmo de tomada de decisão para processamento entre camadas da rede de comunicação.

Nota-se o crescimento dos estudos que investigam os desafios em ambientes HSH, apresentando novas técnicas para monitorar e auxiliar os usuários no ambiente residencial e na qualidade de vida. No entanto, há a falta de estudos que abordem a análise dos custos computacionais incorridos na identificação e monitoramento do cotidiano do usuário, principalmente no

Quadro 3 – Análise dos trabalhos identificados na literatura referente ambientes HSH.

Trabalhos relacionados	Ambiente HSH	Análise emocional	Análise física	Comunicação de camadas	Tomada de decisão
Abowd <i>et al.</i> (2002)	✓	-	-	-	-
Helal <i>et al.</i> (2005)	✓	-	-	-	-
Fernández-Caballero <i>et al.</i> (2014)	✓	✓	X	X	X
Mano <i>et al.</i> (2016a)	✓	✓	X	✓	X
Castillo <i>et al.</i> (2016)	✓	✓	X	X	X
Mohamed <i>et al.</i> (2017)	✓	X	✓	✓	X
Patel, Ali e Sheth (2017)	X	X	X	✓	✓
Kang, Kim e Choo (2017)	X	X	X	✓	✓
Augusto <i>et al.</i> (2018)	X	X	X	✓	✓
Neto <i>et al.</i> (2019)	✓	X	✓	✓	✓
Proposta	✓	✓	✓	✓	✓

que diz respeito aos aspectos emocionais. Nesse sentido, este estudo busca preencher essa lacuna na literatura, apresentando uma arquitetura dinâmica para o monitoramento tanto físico quanto emocional dos usuários em ambientes HSH. A arquitetura estabelecida é baseada em diferentes níveis de processamento computacional; em outras palavras, o carregamento das solicitações de serviço de classificação física e emocional aumenta a carga de trabalho, que é transferida para um nível da arquitetura com maior poder de processamento. Além disso, propomos um algoritmo de tomada de decisão para a comunicação e processamento de informações entre os níveis da arquitetura, balanceando o processamento dependendo da necessidade da carga de trabalho.

3.4 Considerações finais

Este capítulo apresentou uma investigação bibliográfica que, além de investigar a evolução da literatura na área de reconhecimento de emoções, classificação de aspectos físicos e ambientes HSH, teve como objetivo identificar as lacunas para esta pesquisa.

Notou-se, que para o reconhecimento de emoções, apesar do crescimento de estudos que abordam esse tema, não há um consenso referente ao componente emocional e a abordagem de classificação adequada para a classificação de aspectos emocionais. No entanto, de acordo com a teoria que norteia as emoções (Seção 2.2), a abordagem multimodal se mostra uma promissora solução para o reconhecimento das emoções. Ainda, em relação a identificação de aspectos físicos, apesar de trabalhos abordarem das atividades e uma possível queda de pessoas idosas, poucos estudos relacionam questões relacionadas a qualidade do ambiente para o repouso, visto que este fator pode influenciar diretamente na qualidade de vida e bem estar de usuários com idade elevada. Por fim, observou-se a lacuna na literatura referente a uma arquitetura dinâmica para o monitoramento, tanto de aspectos físicos quanto emocionais, dos usuários em ambientes HSH. Destaca-se que a questão de classificação de aspectos físicos e emocionais em ambientes inteligentes é um grande desafio que o grupo de pesquisa vem investigando em seus projetos (MANO *et al.*, 2015; MANO *et al.*, 2016a; ALVES, 2016; MANO; VASCONCELOS;

UEYAMA, 2016; MANO *et al.*, 2016b; MANO; SAWADA; UYAMA, 2016; Pereira Rocha Filho *et al.*, 2018; NETO *et al.*, 2019; NETO *et al.*, 2018; Mano, 2018; SAWADA *et al.*, 2019; MANO *et al.*, 2019a; MANO *et al.*, 2019b; MANO *et al.*, 2019) que também é objeto de pesquisa desta Tese.

INTELIGÊNCIA COMPUTACIONAL PARA IDENTIFICAÇÃO DE ASPECTOS EMOCIONAIS

Neste capítulo, é discutido o uso de inteligência computacional para a identificação de aspectos emocionais. Inicialmente, apresentamos a abordagem para o mapeamento facial e discutimos os resultados do reconhecimento do indivíduo e classificação da emoção pela face. Em seguida, abordamos a classificação da emoção pelo tom de voz e pela descrição da fala, bem como apresentamos a avaliação e discussão dos resultados da classificação de ambas as abordagens. Posteriormente, discutimos a abordagem multimodal para a classificação da emoção, com base na face, tom de voz e descrição da fala, e apresentamos a avaliação e discussão dos resultados. Em complemento, apresentamos um estudo do uso da abordagem multimodal para classificar emoções em um ambiente real. O ambiente selecionado foi o ambiente educacional (simulação clínica) para alunos da área da saúde (enfermagem). Finalmente, é apresentada as considerações finais deste capítulo.

4.1 Identificação do usuário e classificação da emoção pela face

As Expressões Motoras, também conhecidas como Reações Expressivas, são responsáveis pelo comportamento comunicativo do indivíduo ([SCHERER, 2005](#); [MAHLKE; MINGE, 2008](#); [ALVES, 2016](#)). Segundo [Desmet \(2003\)](#), cada emoção está associada a um determinado padrão da face, como por exemplo, a “raiva está associado ao olhar fixo, sobrancelhas contraídas, lábios comprimidos, movimentos vigorosos e rápidos”. As expressões motoras estão relacionadas a expressões faciais, a gestos corporais e a características do discurso da fala ([DESMET, 2003](#); [ALVES, 2016](#)). Complementando a ideia, [Mahlke e Minge \(2008\)](#) sugerem que existe uma

relação entre a dilatação da pupila e o prazer de um estímulo, sendo que estímulos agradáveis são suscetíveis a pupilas mais dilatadas.

O reconhecimento emocional pela face implica em alterações que acompanham a experiência emocional do usuário (ALVES, 2016). O exemplo mais comum são as seis emoções básicas (alegria, desgosto, medo, raiva, surpresa e tristeza) e o conjunto de expressões faciais relacionadas a elas (EKMAN, 1973). Todas as outras categorias emocionais são, então, construídas a partir de combinações dessas emoções básicas (ALVES, 2016; MANO; SAWADA; UHEYAMA, 2016). Esta abordagem é apoiada pelos estudos transculturais, os quais indicam que os seres humanos percebem certas emoções básicas transmitidas por expressões faciais da mesma forma, independentemente da cultura, da etnia e dos costumes (EKMAN, 1973).

No mesmo sentido, a identificação do indivíduo pelo reconhecimento facial é uma das áreas com maior crescimento e influência na biometria moderna (GALTERIO; SHAVIT; HAYAJNEH, 2018). Um dos fatores desse crescente avanço é a natureza não invasiva do método de aquisição dos dados, quando comparado com impressões digitais ou reconhecimento de íris. Diversas são as aplicações que utilizam desse método de identificação de indivíduo: ciência forense, localização de pessoas desaparecidas, verificação para controle de acesso, passaporte, carteira de motorista, sistemas bancários, entre outros (GALTERIO; SHAVIT; HAYAJNEH, 2018).

Diante desse panorama, estudos atuais evidenciam a importância de se considerar as preferências de cada indivíduo, em especial relacionado ao aspecto emocional, provendo uma melhor satisfação na interação homem-máquina (GONÇALVES; UHEYAMA, 2012; ALVES, 2016). O principal objetivo de se promover essa interação é contribuir para o aumento da coerência, da consistência e da credibilidade das reações e das respostas computacionais providas durante a interação humana, via interface humano-computador (ALVES, 2016). Assim, sistemas capazes de reconhecer o usuário e suas emoções, interpretando e reagindo de modo inteligente, poderiam atender mais adequadamente aos requisitos de interação dos indivíduos.

4.1.1 Método para a análise facial - Características geométricas

A abordagem para a identificação do rosto e para a análise da expressão facial tratada neste estudo é baseada em características geométricas (MANO *et al.*, 2015; ALVES, 2016; Mano, 2018). A face é utilizada na modelagem para uma abordagem que se assemelhe à maneira como os seres humanos interpretam os elementos da face. Diferentes representações faciais (alegria, desgosto, medo, neutro, raiva, surpresa e tristeza) são propostas para a identificação das emoções e identificação do usuário, codificando a configuração facial de um indivíduo. A Figura 4 ilustra o processo de mapeamento para o reconhecimento e tratamento da face, para o reconhecimento do indivíduo e classificação da emoção.

Nesse caso, é feito um processo de mapeamento da face considerando 33 pontos obtidos

Figura 4 – Processo de mapeamento da face.

Fonte: Mano *et al.* (2015), Alves (2016), Mano (2018).

na extração de características faciais (MANO *et al.*, 2015; ALVES, 2016; Mano, 2018). Os pontos são partes específicas da face que, quando juntas, identifica de maneira única um indivíduo. Os pontos são distribuídos da seguinte maneira: i.) oito para mapear a boca; ii.) seis para cada um dos olhos; iii.) três para cada sobrancelha; iv.) três para o queixo; v.) dois para as narinas e; vi.) dois para demarcar as extremidades laterais da face perto dos olhos. Além disso, as distâncias e os ângulos são obtidos para todas as combinações possíveis de pontos. Tal modelagem cria uma representação com dimensionalidade de $D_1 = 33 \times 33 + 8 + 33 = 1.130$ atributos (MANO *et al.*, 2015; ALVES, 2016; Mano, 2018).

4.1.2 Reconhecimento do indivíduo - Avaliação e discussão

Para os experimentos de identificação do usuário, foi adotada uma abordagem que é semelhante à forma como os seres humanos interpretam as características do rosto, isto é, diversas fotos do mesmo indivíduo foram utilizadas para o modelo de identificação do usuário. Diferentes representações faciais (neutro, feliz, triste, com medo, irritado e surpreso) do mesmo indivíduo são usadas para identificar a pessoa; o uso de múltiplas imagens faciais torna possível codificar uma configuração única para cada indivíduo. O subconjunto selecionado da base de dados *Extended Cohn-Kanade* (CK+) (LUCEY; COHN *et al.*, 2010) inclui todas as expressões existentes de diferentes atores. Para o experimento conduzido para identificação do usuário, foram selecionados 20 indivíduos diferentes, com 100 imagens cada, perfazendo um **total de 2.000 imagens**, a fim de atingir um bom grau de precisão.

Inicialmente, analisamos o desempenho dos algoritmos através da técnica de Planejamento Experimental e Avaliação; neste caso, empregamos a validação cruzada *k-fold cross-validation* com $k = 10$, com $k - 1$ para treinamento e o restante para teste. Assim, é possível medir a estimativa de erro com mais precisão, pois a estimativa do valor médio tende a uma taxa de erro zero real à medida que aumenta o n , o que geralmente é o caso de pequenos conjuntos de

exemplos (ALVES, 2016; Mano, 2018). Desta forma, foi possível obter a taxa média de precisão para cada algoritmo. A Figura 5 apresenta os Boxplots¹ referentes às classificações realizadas pelos algoritmos selecionados e a Tabela 1 mostra a acurácia média na identificação dos usuários.

Figura 5 – Boxplots das acurácias apresentadas pelos classificadores para a identificação do indivíduo.

Fonte: Elaborada pelo autor.

Tabela 1 – Média (%) das acurácias e os p -valores dos conjuntos de resultados para a identificação de indivíduo.

Classificadores	Acurácia (%)	p -valores (<i>Shapiro Wilk</i>)
Árv. Dec	94,00	0.4848
Fuzzy	83,00	0.6292
kNN	99,00	0.0068
N. Bayes	97,95	0.8746
SVM	99,75	0.0020

Posteriormente, realizamos análises estatísticas para validar os resultados. Inicialmente, utilizou-se o método de *Shapiro-Wilk*² para verificar sua adequação à hipótese de normalidade e, portanto, indicar para testes paramétricos ou não paramétricos. Como nem todos os p -valores

¹ O Boxplot, ou diagrama de caixa, é um gráfico que capta importantes aspectos de um conjunto de dados através do seu resumo dos cinco números, formado pelos seguintes valores: valor mínimo, primeiro quartil, segundo quartil, terceiro quartil e valor máximo.

² O teste de *Shapiro-Wilk* é um teste geral projetado para detectar todas as desvios a normalidade. O teste rejeita a hipótese de normalidade quando o valor p é menor ou igual a 0.05. Falhando o teste de normalidade permite afirmar com 95% de confiança que os dados não se encaixam a distribuição normal. Passar no teste de normalidade só lhe permite declarar que nenhum desvio significativo à normalidade foi encontrado.

obtidos são maiores que 0.05 (veja Tabela 1), rejeitamos a hipótese de normalidade com 95% de confiabilidade. Isso significa que o teste não paramétrico é o mais adequado para a próxima análise.

Para o teste não paramétrico, realizamos as comparações de pares com o teste *Wilcoxon Rank Sum*³, no intuito de verificar a diferença estatisticamente significativa entre os grupos de resultados (ver Tabela 2). O teste afirma que valores inferiores a 0.05 indicam diferença estatisticamente significativa entre os grupos de resultados.

Tabela 2 – P-valores da comparação de pares realizada com o teste *Wilcoxon Rank Sum* para a identificação do indivíduo.

	Árv. Dec.	Fuzzy	kNN	N. Bayes
Fuzzy	0.0014	-	-	-
kNN	0.0014	0.0014	-	-
N. Bayes	0.0014	0.0025	0.0014	-
SVM	0.0014	0.0014	0.0031	0.0014

Os resultados demonstram que os algoritmos apresentam diferença estatisticamente significativa entre eles, visto que todas as comparações entre os pares de resultados são inferiores à 0.05. Nesse sentido, podemos afirmar que, para o conjunto de dados utilizados para a identificação do indivíduo, o algoritmo SVM supera as outras soluções. Embora o algoritmo kNN tenha alta taxa de precisão, com 99%, o SVM teve uma taxa de precisão de 99,75% nas experiências realizadas.

4.1.3 Classificação da emoção - Avaliação e discussão

Para os experimentos de classificação emocional por meio da face, foram utilizadas as bases de dados *Radboud Faces* (RaFD) (LANGNER *et al.*, 2010), CK+ (LUCEY; COHN *et al.*, 2010), IMPA-FACE3D (MENA-CHALCO; CESAR-JR; VELHO, 2008) e FACES (EBNER; RIEDIGER; LINDENBERGER, 2010), todas de acesso livre. O Quadro 4 apresenta uma descrição das bases utilizadas, incluindo o número de atores para cada expressão facial avaliada. Para o experimento foram utilizadas um total de **3.115 imagens**, divididas em 445 imagens para cada emoção (alegria, desgosto, medo, neutro, raiva, surpresa e tristeza), de modo a alcançar um balanceamento ideal para os testes realizados.

Inicialmente, analisamos o desempenho dos algoritmos com a técnica de Planejamento Experimental e Avaliação *k-fold cross-validation* com $k = 10$, com $k - 1$ para treinamento e o

³ O teste de *Wilcoxon*, que se refere ao teste de *Rank Sum* ou o teste de classificação assinado, é um teste não paramétrico que compara dois grupos emparelhados. O teste essencialmente calcula a diferença entre cada conjunto de pares e analisa estas diferenças. O teste de *Wilcoxon Rank Sum* pode ser utilizado para testar a hipótese nula de que duas populações têm a mesma distribuição contínua.

Quadro 4 – Descrição das bases de dados para a classificação emocional da face.

Base	Descrição	Nº de indivíduos
RaFD	Expressões faciais de adultos e crianças, do sexo masculino e feminino, de nacionalidades americana e marroquina	67
CK+	Expressões faciais de atores adultos, 69% mulheres e 31% homens. Sendo, 81% europeus e americanos, 13% afro-americanos e 6% pertencentes a outros grupos étnicos.	123
IMPA-FACE3D	Composto por 22 homens e 16 mulheres, com a maioria dos indivíduos com idade entre 20 e 50 anos	38
FACES	Divididos em 58 jovens, 56 pessoas de meia-idade e 56 idosos, entre homens e mulheres	171

restante para teste. Os resultados demonstraram que a abordagem SVM fornece uma classificação mais precisa do que o uso dos outros classificadores selecionados. Isso pode ser visto na Figura 6, que mostra os Boxplots para cada tipo de classificador empregado. O Boxplot com cor destacada refere-se aos resultados do SVM e exibe a mediana da acurácia superior se comparada às alcançadas pelos demais classificadores (os valores das medianas podem ser encontrados na Tabela 3). Além disso, uma menor dispersão dos resultados obtidos pelo SVM pode ser detectada, o que mostra uma maior estabilidade em sua execução.

Figura 6 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção pela face.

Fonte: Elaborada pelo autor.

Posteriormente, utilizamos o método de *Shapiro-Wilk* para verificar sua adequação à hipótese de normalidade e, portanto, indicar para testes paramétricos ou não paramétricos. Como todos os p -valores obtidos são maiores que 0.05 (veja Tabela 3), aceitamos a hipótese de normalidade com 95% de confiabilidade. Isso significa que o teste paramétrico é o mais adequado para a próxima análise.

Tabela 3 – Média (%) das acurácias e os p -valores dos conjuntos de resultados para a classificação da emoção.

Classificadores	Acurácia	p -valores (<i>Shapiro Wilk</i>)
Árv. Dec.	68,10	0.1317
Fuzzy	65,20	0.6184
kNN	69,47	0.2057
N. Bayes	70,79	0.5028
SVM	76,16	0.4378

Para o teste paramétrico, utilizamos o teste- T ⁴, método paramétrico que permite uma confiabilidade maior na análise de dados com adequação à distribuição normal. As comparações pareadas feitas com o teste- T são mostradas na Tabela 4 e valores inferiores a 0.05 indicam diferença estatisticamente significativa entre os grupos de resultados.

Tabela 4 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela face.

	Árv. Dec.	Fuzzy	kNN	N. Bayes
Fuzzy	0.08615	-	-	-
kNN	0.53537	0.00525	-	-
N. Bayes	0.09942	0.00026	0.53537	-
SVM	3.6e-07	1.3e-10	1.5e-05	0.00039

Os p -valores indicam que o SVM apresenta diferença estatisticamente significativa em relação aos outros classificadores e os resultados demonstram que, de todos os classificadores individuais empregados, o SVM é a técnica que alcança maior precisão na avaliação emocional (ver Tabela 3). Finalmente, a Figura 7 mostra a Matriz de Confusão⁵, na qual as colunas representam os dados de referência, enquanto as linhas representam a classificação gerada pelo SVM. Vale ressaltar, que de todas as emoções avaliadas, o modelo fornece uma melhor classificação das emoções “Alegria”, “Desgosto” e “Surpresa”.

⁴ O teste- T é um teste de hipótese que utiliza conceitos estatísticos para rejeitar ou não uma hipótese nula. Essa premissa é normalmente utilizada quando a estatística de teste segue uma distribuição normal.

⁵ Matriz de Confusão é um *layout* de tabela específico que permite a visualização do desempenho de um algoritmo, tipicamente um aprendizado supervisionado.

Figura 7 – Matriz de Confusão da classificação da face para emoção (abordagem SVM).

	Alegria	Desgosto	Medo	Neutro	Raiva	Surpresa	Tristeza
Alegria	406	13	8	2	5	6	5
Desgosto	21	340	9	14	34	5	22
Medo	8	16	310	41	14	17	39
Neutro	4	12	38	301	18	9	63
Raiva	8	54	7	25	312	2	37
Surpresa	4	5	18	2	9	403	4
Tristeza	8	18	27	51	36	5	300

Fonte: Elaborada pelo autor.

4.2 Classificação da emoção pelo tom de voz

O sinal da fala é constituído por diversos componentes de frequência (ver Figura 8) e varia de acordo com o teor emocional da elocução produzida pelo usuário. De maneira mais específica, a utilização da fala dos usuários para identificar as emoções permite que tal informação possa ser obtida sem uma ação explícita do usuário e sem o uso de equipamentos específicos ou intrusivos (MANO; VASCONCELOS; UEYAMA, 2016).

Figura 8 – Representação dos componentes de frequência da fala.

Fonte: Mano, Vasconcelos e Ueyama (2016).

A diferença entre sinais da voz que expressam diferentes emoções está, sobretudo, na forma como o espectro de energia do sinal se comporta (MANO; VASCONCELOS; UEYAMA, 2016). Por exemplo, emoções cuja expressão vocal é caracterizada por exaltações e brados, tendem a carregar mais energia do que emoções cuja expressão oral é mais branda. Assim, ao compararmos o espectro de frequências de uma frase pronunciada de maneira negativa ao espectro da mesma frase pronunciada de maneira positiva, notaremos diferenças na distribuição espectral de energia (MANO; VASCONCELOS; UEYAMA, 2016), como ilustra a Figura 9.

Figura 9 – Espectrograma da frase “Gli operari si alzarò presto” pronunciada de maneira negativa e positiva no idioma italiano.

Fonte: Mano, Vasconcelos e Ueyama (2016).

4.2.1 Método para análise da voz - Espectro de energia

A elocução humana é produto da filtragem realizada pelo trato vocal sobre a onda sonora resultante da interação do fluxo de ar vindo dos pulmões com as cordas vocais (FURUI, 2018; NETO *et al.*, 2018; SAWADA *et al.*, 2019). A maneira como uma frase é pronunciada pode levar a inúmeras interpretações, e seres humanos são capazes de atribuir conotações emocionais a elocuições de maneira automática (MANO; VASCONCELOS; UEYAMA, 2016). Do ponto de vista auditivo, as múltiplas interpretações possíveis são fruto de como a energia do sinal da fala está distribuída sobre o espectro de frequências audíveis do sinal, em outras palavras, da densidade espectral de energia (MANO; VASCONCELOS; UEYAMA, 2016). Características associadas à densidade espectral de energia de sinais sonoros são o “padrão” em aplicações de reconhecimento de voz, de instrumentos e de estilos musicais. Estudos (MARTIN; JURAFSKY, 2009; MANO; VASCONCELOS; UEYAMA, 2016; NETO *et al.*, 2018; SAWADA *et al.*, 2019) apontam *Mel Frequency Cepstral Coefficient* (MFCC) como característica eficiente para o processamento da fala. Esse tipo de *feature* é baseado em cepstro e inspirado na maneira como o ouvido humano responde a estímulos sonoros, uma vez que as frequências do espectro estão em escala mel, não linear, com gradação que procura imitar a percepção auditiva humana (MANO; VASCONCELOS; UEYAMA, 2016).

Diferentes representações vocais (alegria, desgosto, medo, neutro, raiva, surpresa e tristeza) são propostas para a identificação e a classificação das emoções capazes de codificar a configuração vocal do indivíduo. A ferramenta utilizada para obter as informações vocais é o openSMILE (EYBEN *et al.*, 2013), baseado em *scripts* de configuração da extração que se deseja executar. A ferramenta recebe como argumentos o caminho do *script* de extração a ser utilizado e o caminho do arquivo de áudio do qual são extraídas as *features* descritas. O *pipe* de extração do openSMILE é apresentado na Figura 10. O processo de extração de MFCCs (representado por Melspec) e de *Log Energy* (representado por LOGE) são idênticos até o janelamento do sinal e se divergem em seguida. Os processos convergem novamente para extrair os coeficientes delta e aceleração (MANO; VASCONCELOS; UEYAMA, 2016).

Figura 10 – Representação do *pipe* de extração de *features* da voz.

Fonte: Mano, Vasconcelos e Ueyama (2016).

Neste estudo, 12 MFCCs, *features log energy*, delta e aceleração foram extraídos para formarem um conjunto de 39 *features*. O coeficiente *log energy* é obtido a partir do cômputo do logaritmo da energia (total) do fragmento de sinal digital $x[n]$ (ver Equação 4.1). As *features* delta e aceleração são respectivamente as derivadas de primeira e de segunda ordem do vetor inicial de 12 MFCCs + 1 coeficiente *log energy*.

$$E(x) = \sum_{n=0}^{N-1} x[n]^2 \quad (4.1)$$

4.2.2 Avaliação e discussão

Para os experimentos de classificação emocional por meio da voz, isto é, de acordo com o espectro de energia da locução, foram utilizados as bases de dados EMOVO (COSTANTINI *et al.*, 2014) e VERBO (NETO *et al.*, 2018) compostas de atores, atribuindo-lhes conteúdo emocional para pronunciar frases, ambas de acesso livre. O Quadro 5 apresenta uma descrição das bases utilizadas, incluindo o número de atores e o número de frases para cada locução emocional avaliada. Para o experimento foram utilizadas um total de **1.757 áudios**, divididas em 251 áudios para cada emoção (alegria, desgosto, medo, neutro, raiva, surpresa e tristeza), de modo a alcançar um balanceamento ideal para os testes realizados.

Quadro 5 – Descrição das bases de dados para a classificação emocional da voz.

Base	Descrição	Nº de indivíduos	Nº de frases
EMOVO	atores italianos constituem a base, 3 mulheres e 3 homens, com idades entre 23 a 30 anos	6	98
VERBO	atores brasileiros constituem a base, 6 mulheres e 6 homens, entre 2 e 23 anos de experiência	12	167

Como nos experimentos da análise emocional pela face, inicialmente, analisamos o desempenho dos algoritmos por meio da técnica de Planejamento Experimental e Avaliação

k-fold cross-validation com $k = 10$, com $k - 1$ para treinamento e o restante para teste, obtendo a taxa média de precisão para cada algoritmo. Os resultados demonstraram que a abordagem do kNN fornece uma classificação mais precisa do que o uso dos outros classificadores selecionados. Isso pode ser visto na Figura 11, que mostra os Boxplots para cada tipo de classificador empregado. O Boxplot com cor destacada refere-se aos resultados do kNN e exibe a mediana da acurácia superior se comparada aos demais classificadores (os valores das medianas podem ser encontrados na Tabela 5).

Figura 11 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção do espectro de energia da voz.

Fonte: Elaborada pelo autor.

Tabela 5 – Média (%) das acurácias e os p -valores dos conjuntos de resultados da classificação da voz.

Classificadores	Acurácia	p -valores (<i>Shapiro Wilk</i>)
Árv. Dec.	67,79	0.1770
Fuzzy	65,47	0.5453
kNN	74,90	0.3951
N. Bayes	66,62	0.2118
SVM	68,80	0.2681

Seguindo a mesma condição anterior, realizamos análises estatísticas para validar os resultados. Inicialmente, utilizou-se o método de *Shapiro-Wilk* para verificar sua adequação à hipótese de normalidade e, portanto, indicar a testes paramétricos ou não paramétricos. Como

todos os p -valores obtidos são maiores que 0.05 (veja Tabela 5), aceitamos a hipótese de normalidade e seguimos com o teste- T como teste paramétrico para a próxima análise (Tabela 6).

Tabela 6 – P -valores da comparação de pares realizada com o teste- T para a classificação da emoção pela voz.

	Árv. Dec.	Fuzzy	kNN	N. Bayes
Fuzzy	0.38449	-	-	-
kNN	1.1e-05	2.9e-08	-	-
N. Bayes	1.0000	1.00000	5.7e-07	-
SVM	1.0000	0.07471	0.00014	0.38449

Os p -valores da comparação de pares realizada com o teste- T indicam que o kNN apresenta diferença estatisticamente significativa em relação aos outros classificadores e os resultados demonstram que, de todos os classificadores individuais empregados, o kNN é a técnica que alcança maior precisão na avaliação emocional pelo tom de voz (ver Tabela 5). Finalmente, a Figura 12 mostra a Matriz de Confusão gerada pelo kNN na classificação da emoção pelo tom da voz.

Figura 12 – Matriz de Confusão da classificação da voz para emoção (abordagem kNN).

Fonte: Elaborada pelo autor.

Diferente dos resultados da classificação da face (Seção 4.1.3), o modelo para a classificação do tom de voz apresenta um comportamento distinto, com um maior grau de precisão para as emoções com alto, baixo ou mesmo nenhum grau de energia na maneira da pronuncia da frase, isto é, as emoções “Neutro”, “Raiva” e “Alegria”.

4.3 Classificação da emoção pela descrição da fala

As pessoas geralmente diferem consideravelmente umas das outras, em seus padrões de pensamento, bem como sentimentos ou ações. Essas diferenças se refletem no que dizem, sobre o que pensam, sentem ou fazem. Pesquisadores de psicolinguística consideram as palavras usadas na vida cotidiana como um reflexo de quem somos, como nos sentimos e como pensamos (SAWADA *et al.*, 2019). Estudos encontraram ligações sistemáticas entre personalidade e uso da linguagem em uma ampla gama de diferentes contextos, incluindo atribuições de escrita direta (PENNEBAKER; KING, 1999; HIRSH; PETERSON, 2009), entrevistas estruturadas (FAST; FUNDER, 2008) e gravações naturalistas da vida cotidiana (MEHL; GOSLING; PENNEBAKER, 2006).

A frequência com que utilizamos certos tipos de palavras pode fornecer pistas que podem ajudar a inferir traços de personalidade, estilos de pensamento, conexões sociais e estados emocionais (SAWADA *et al.*, 2019). Pennebaker e King (1999) e Lee *et al.* (2007) realizaram correlações que são teoricamente previsíveis entre as dimensões de Extroversão e Neuroticismo e o uso de palavras que denotam categorias de emoções positivas e negativas. Esse tipo de pesquisa faz uma correlação significativa entre a escolha de palavras e tipos de personalidade, emoções, atitudes, necessidades, valores e processos de pensamento, bem como a relação de traços de personalidade com a frequência com que certas categorias de palavras são usadas (FAST; FUNDER, 2008; HIRSH; PETERSON, 2009; YARKONI, 2010).

A maioria das palavras definidas neste contexto é baseada no dicionário chamado *Linguistic Inquiry and Word Count* (LIWC) (TAUSCZIK; PENNEBAKER, 2010), usado com o propósito de encontrar categorias psicologicamente significativas, baseadas no uso de palavras escritas. A estratégia analítica tradicional que emprega o Processamento de Linguagem Natural (PLN) estabelece uma base de compreensão para o seu discurso através do seu vocabulário, mesmo quando aplicada a uma análise linguística superficial, como a contagem de frequência, sendo uma ferramenta suplementar eficaz para analisar o comportamento da linguagem.

4.3.1 Método para análise da descrição da fala

O processamento da descrição da fala foi dividida em três etapas distintas: i.) transcrição da fala, ii.) normalização e iii.) transformação. A Figura 13 ilustra as etapas para o processamento da descrição da fala, baseada na técnica *Bag of Words*.

Na etapa de transcrição da fala, visto que as sentenças estão em formato de áudio, utilizamos a API *Speech to Text* (IBM, 2018a) para transformar a fala contida no áudio em um documento de texto. Para a etapa de normalização foram aplicadas técnicas de PLN, por meio da biblioteca *Natural Language Toolkit* (NLTK), para “limpeza do texto”, isto é, foram retirados sentenças indesejadas que não são significativas, extração de *tokens* e remoção de *stops words*. Por fim, a etapa de transformação é utilizada para converter o texto em características

Figura 13 – Etapas para o processamento da descrição da fala.

Fonte: Elaborada pelo autor.

mais significativas para os algoritmos de ML. A transformação do texto é realizada por meio da técnica conhecida como *Bag of Words*. Essa técnica transforma a coleção de textos em uma tabela, que indica as palavras e o número de ocorrência dos termos dentro do corpo textual, assim o valor textual é transformado em numérico para ser utilizado pelos algoritmos de classificação (AGUIAR *et al.*, 2018).

4.3.2 Avaliação e discussão

Para os experimentos de classificação emocional por meio da descrição da fala selecionamos a base de dados proposto por Aman e Szpakowicz (2007). A base foi selecionada por apresentar as seis emoções básicas proposta por Ekman (2006), juntamente com a emoção neutro, como tratado nos experimentos anteriores. O conjunto de dados contém 4.090 frases anotadas de postagens de *blog*, sendo 68% das quais foram marcadas como neutra. Visto o alto desequilíbrio da base de dados, somente um subconjunto foi selecionado, buscando um equilíbrio ideal para o experimento. Sendo assim, **805 sentenças** emocionais foram utilizadas, divididas em 115 para cada estado emocional.

Como nos experimentos anteriores, analisamos o desempenho dos algoritmos por meio da técnica de Planejamento Experimental e Avaliação *k-fold cross-validation* com $k = 10$, com $k - 1$ para treinamento e o restante para teste. Os resultados demonstraram que a abordagem N. Bayes fornece uma classificação mais precisa do que o uso dos outros classificadores. Isso pode

ser visto na Figura 14, que mostra os Boxplots para cada tipo de classificador empregado. O Boxplot com cor destacada refere-se aos resultados do N. Bayes e exibe a mediana da acurácia superior (os valores das medianas podem ser encontrados na Tabela 7).

Figura 14 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção mediante a descrição da fala.

Fonte: Elaborada pelo autor.

Tabela 7 – Média (%) das acurácias e os p -valores dos conjuntos de resultados da classificação da descrição da fala.

Classificadores	Acurácia	p -valores (<i>Shapiro Wilk</i>)
Árv. Dec.	67,49	0.5510
Fuzzy	65,57	0.2910
kNN	66,38	0.6290
N. Bayes	72,47	0.2057
SVM	67,17	0.1542

Posteriormente, utilizou-se o método de *Shapiro-Wilk* para verificar sua adequação à hipótese de normalidade e, portanto, indicar para testes paramétricos ou não paramétricos. Como todos os p -valores obtidos são maiores que 0.05 (veja Tabela 7), aceitamos a hipótese de normalidade e seguimos com o teste- T como teste paramétrico para a próxima análise (Tabela 8).

Os p -valores da comparação de pares realizada com o teste- T indicam que o N. Bayes apresenta diferença estatisticamente significativa em relação aos outros classificadores e os resultados demonstram que, de todos os classificadores individuais empregados, o N. Bayes é a

Tabela 8 – *P*-valores da comparação de pares realizada com o teste-*T* para a classificação da emoção pela descrição da fala.

	Árv. Dec.	Fuzzy	kNN	N. Bayes
Fuzzy	0.57141	-	-	-
kNN	1.0000	1.0000	-	-
N. Bayes	0.00043	2.0e-06	2.2e-05	-
SVM	1.0000	0.81673	1.0000	0.00019

técnica que alcança maior precisão na avaliação emocional pela descrição da fala (ver Tabela 7). Finalmente, a Figura 15 mostra a Matriz de Confusão gerada pelo N. Bayes na classificação da emoção pela descrição da fala. Vale ressaltar que o modelo baseado no N. Bayes fornece uma melhor classificação das emoções “Alegria”, “Raiva”, “Surpresa” e “Tristeza”.

Figura 15 – Matriz de Confusão da classificação da voz da descrição da fala (abordagem N. Bayes).

Fonte: Elaborada pelo autor.

4.4 Abordagem multimodal para a classificação da emoção

Diante o desafio de classificar e determinar as emoções, a relação multimodal de diferentes componentes são estudados em pesquisas sobre o reconhecimento das emoções (PORIA *et al.*, 2017; GONÇALVES *et al.*, 2017a; FIORI; VESELY-MAILLEFER, 2018). A abordagem multimodal oferece uma maneira mais abrangente para entender os diferentes aspectos das emoções em todos os tipos de situações (MAHLKE; MINGE, 2008; GONÇALVES *et al.*, 2017a)

e podem ser utilizados para a análise emocional, sem desconsiderar a relação entre os diferentes componentes (SCHERER, 2005).

O aprendizado multimodal é aquele que considera informações de múltiplas fontes para realizar uma determinada tarefa (PORIA *et al.*, 2017; GONÇALVES *et al.*, 2017a). Essa hipótese é amparada como a informação é processada no mundo real, onde as informações se complementam para formalizar um dado conceito na mente humana. Por exemplo, imagens são relacionadas a legendas e palavras-chaves, notícias são complementadas com imagens, vídeos e áudio (PORIA *et al.*, 2017).

De fato, o objetivo do conceito de aprendizado multimodal é aumentar a precisão e confiabilidade das estimativas (PORIA *et al.*, 2017). No entanto, definir os componentes e métodos para a classificação emocional ainda é uma tarefa difícil, visto a dificuldade de selecionar os componentes ideais e métodos de classificação que possam inferir a emoção humana de maneira mais assertiva e, um dos principais obstáculos à fusão multimodal é o desenvolvimento e especificação de uma metodologia para integrar informações cognitivas e afetivas de diferentes fontes, em diferentes escalas de tempo e valores de medição.

4.4.1 Metodologia multimodal para a emoção

Como discutido na Seção 4.4, existem algumas técnicas principais para a fusão multimodal de dados: fusão em nível de característica, fusão em nível de decisão, fusão híbrida e fusão em nível de modelo. Nós implementamos a fusão em nível de característica, utilizando os recursos de três modalidades distintas (face, tom de voz e descrição da fala) como descritas anteriormente nas Seções 4.1, 4.2 e 4.3, respectivamente. Esta metodologia foi então utilizada para propor uma estrutura, na qual múltiplos sensores são empregados, para avaliar as emoções de um usuário no momento da interação, como ilustrado na Figura 16.

A estrutura utiliza o modelo de classificação de ML para combinar as saídas do processamento de características e definir o estado emocional do usuário. À medida que o aprendizado supervisionado é conduzido, o método usa instâncias rotuladas para aprender sobre os modelos.

4.4.2 Avaliação e discussão

Para avaliar nosso modelo, utilizamos a base de dados One-Minute-Gradual Emotion Dataset (OMG-Emotion Dataset) (BARROS *et al.*, 2018a) composto por vídeos de emoção relativamente longos, com duração média de 1 minuto, coletados de uma variedade de canais do YouTube. Os vídeos foram selecionados utilizando uma técnica de rastreamento que usa palavras-chave específicas baseadas em comportamentos emocionais de longo prazo, como monólogos, audições, diálogos e cenas emocionais. O uso de tais vídeos permitiu que diferentes comportamentos emocionais fossem apresentados em um contexto que mudou gradualmente com o tempo. Os vídeos foram anotados por pelo menos cinco sujeitos independentes usando a

Figura 16 – Estrutura para a classificação multimodal da emoção baseada na face, tom de voz e descrição da fala.

Fonte: Elaborada pelo autor.

ferramenta *Amazon Mechanical Turk*. Para manter as informações contextuais de cada vídeo, cada sujeito observou os vídeos utilizando uma escala de excitação/valência. Vale salientar, que cada avaliador recebeu as informações contextuais do vídeo do conjunto de dados. Isso significa que cada observador levou em consideração não apenas a imagem e a informação do áudio, mas também o contexto de cada vídeo. Desta forma, o OMG-Emotion Dataset é baseado em informações multimodais, apresentando benefícios em pesquisas intermodais, especialmente ao analisar as modalidades de imagem, áudio e linguagem.

A base OMG-Emotion Dataset é composto por um total de 420 vídeos, apresentando as emoções básicas: alegria, desgosto, medo, neutro, raiva, surpresa e tristeza. No entanto, o conjunto de dados para tais emoções é desbalanceado, o que pode prejudicar a eficiência na classificação dos algoritmos. Assim, de modo a alcançar um balanceamento ideal para os testes, a base de dados foi dividida em 140 vídeos para cada estado emocional (positivo, neutro e negativo) de acordo com as emoções dos rótulos dos vídeos, como pode ser visto no Quadro 6.

Quadro 6 – Descrição do balanceamento da base de dados OMG-Emotion Dataset.

Rótulo	Emoções	Nº de vídeos
Positivo	alegria e surpresa	140
Neutro	neutra	140
Negativo	aversão, medo, raiva e tristeza	140

Inicialmente, analisou-se o desempenho dos algoritmos para a classificação da emoção adotando as metodologias para a identificação da imagem, do som e do contexto (como descrito nas Seções 4.1, 4.2 e 4.3, respectivamente). Para os experimentos foi adotado a técnica de Planejamento e Avaliação de Experimentos *k-fold cross-validation* com $k = 10$, sendo $k - 1$ para treinamento e o restante para teste. Os resultados demonstram que cada metodologia (face, voz e

contexto) possibilita uma classificação diferente de acordo com o algoritmo selecionado. Isso pode ser visto na Figura 17, a qual apresenta os Boxplots referentes as classificações realizadas e a Tabela 9 apresenta as medianas alcançadas pelos algoritmos.

Figura 17 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção dos vídeos.

Fonte: Elaborada pelo autor.

Tabela 9 – Acurácia média (%) dos algoritmos avaliados.

Classificadores	Face	Voz	Contexto
N. Bayes	68,78	65,33	72,47
Árv. Dec.	66,90	68,58	67,50
kNN	67,47	71,80	66,40
Fuzzy	63,69	65,17	65,58
SVM	72,31	66,80	67,17

Tais resultados demonstram que, para a metodologia adotada, cada classificador selecionado para os testes possui uma acurácia distinta entre eles. Para a classificação da imagem facial o SVM apresenta desempenho superior, visto que essa abordagem proporciona acurácia média superior (72,31%) e menor dispersão dos resultados, o que demonstra uma maior eficiência e estabilidade em suas execuções. Por outro lado, para a classificação do tom de voz, o classificador que apresentou os melhores resultados foi o kNN com acurácia média de 71,80%. Por fim, para a classificação do contexto do vídeo por meio do que é falado, os melhores resultados foram alcançados pelo N. Bayes com acurácia média de 72,47%. Vale salientar, que os resultados

dos testes acompanharam os resultados apresentados anteriormente para cada metodologia (ver Seções 4.1.3, 4.2.2 e 4.3.2).

A Figura 18 apresenta a Matriz de Confusão dos algoritmos com melhor desempenho referente a classificação dos vídeos para o estado emocional, em que as colunas representam os dados de referência, enquanto as linhas representam a classificação gerada a partir da classificação da emoção.

Figura 18 – Matriz de Confusão referente a classificação da face (SVM), voz (kNN) e contexto (N. Bayes).

Note-se, por meio das Matrizes de Confusão apresentadas (Figura 18), que as classificações geradas pelo classificadores apresentam uma discrepância no que diz respeito a classificação do estado emocional. Enquanto que para a classificação da face (SVM) e do contexto (N. Bayes) os modelos apresentam melhores classificações para os estados emocionais positivo e negativo, a classificação da voz (kNN) apresenta melhor desempenho para o estado neutro. Tais resultados colaboram para a hipótese que a fusão dos sensores pode conduzir a uma melhoria no desempenho em termos de melhor generalização e/ou em termos de aumento da acurácia na classificação da emoção.

Assim, para os testes seguintes, analisou-se o desempenho dos algoritmos de forma conjunta adotando a metodologia de identificação da imagem, da voz e do contexto. Igualmente aos testes anteriores, no sentido de comparação de desempenho e acurácia, foi adotado o *k-fold cross-validation* com $k = 10$, sendo $k - 1$ para treinamento e o restante para teste. A Figura 19 e a Tabela 10 apresentam os Boxplots e a acurácia média referentes as classificações dos algoritmos quando utilizadas as metodologias propostas em conjunto e comparadas as metodologias utilizadas separadamente.

Diante disso, foi utilizado métodos estatísticos para realizar uma análise mais confiável dos dados. Inicialmente é utilizado o método de *Shapiro-Wilk* para avaliar se os grupos de resultados possuem adequação a distribuição normal. Os *p*-valores obtidos são descritos em detalhes na Tabela 10, os quais confirmam a hipótese de adequação a distribuição normal de todos os grupos de resultados, pois todos os valores são maiores que 0.05.

Tal resultado direciona a análise para o uso do teste-*T*, método paramétrico que permite uma confiabilidade maior na análise de dados com adequação à distribuição normal. A Tabela 11

Figura 19 – Boxplots das acurácias apresentadas pelos classificadores para determinar a emoção dos vídeos, levando em consideração a abordagem multimodal.

Tabela 10 – Acurácia média (%) e *p*-valores (*Shapiro Wilk*) dos algoritmos avaliados.

Classificadores	Acurácia	<i>p</i> -valores (<i>Shapiro Wilk</i>)
N. Bayes	78,98	0.7238
Árv. Dec.	77,17	0.5847
kNN	80,17	0.9568
Fuzzy	78,27	0.0939
SVM	83,57	0.7267

apresenta os valores da comparação em pares dos algoritmos utilizados, a fim de checar se existe diferença estatisticamente significativa entre os resultados exibidos pelos classificadores.

Tabela 11 – *P*-valores da comparação de pares realizada com o teste-*T* para a classificação da emoção pela abordagem multimodal.

	N. Bayes	Árv. Dec.	kNN	Fuzzy
Árv. Dec.	0.27867	-	-	-
kNN	0.67795	0.01992	-	-
Fuzzy	0.67795	0.67795	0.27867	-
SVM	0.00017	3.8e-07	0.00706	1.6e-05

O uso do teste-*T* indica que a classificação realizada pelo SVM, além de maior acurácia média, como apresentado na Figura 19 e Tabela 10, também apresenta diferença estatisticamente significativa em relação aos demais classificadores, visto a presença de valores inferiores a 0.05

no teste-*T*. A Figura 20 apresenta a Matriz de Confusão do SVM para a classificação emocional dos vídeos, visto que tal algoritmo obteve melhor desempenho e acurácia comparado aos outros utilizados.

Figura 20 – Matriz de Confusão da classificação multimodal da face, voz e contexto (abordagem SVM).

Fonte: Elaborada pelo autor.

Nota-se, que diferentemente das Matrizes de Confusão apresentadas na Figura 18, onde é visível uma discrepância no que diz respeito a classificação do estado emocional, a Matriz de Confusão dos testes do SVM apresenta uma classificação mais homogênea para a classificação emocional positivo, neutro e negativo dos experimentos realizados com a OMG-Emotion Dataset. Tal cenário reforça a premissa que o uso de metodologias em conjunto para a classificação da emoção é robusto e pode ser escalável para uma classificação mais assertiva do estado emocional do usuário.

4.5 Classificação multimodal da emoção em ambiente real - Simulação clínica

A simulação recorre da criação de uma situação artificial para antecipar a vivência de uma situação real, visando sua compreensão e/ou gestão com o objetivo de aprender, praticar, avaliar, testar ou desenvolver a compreensão dos sistemas ou ações humanas (CENTER, 2013). É um processo de ensino aprendizagem significativo, no qual o conhecimento prévio do aprendiz incorpora conceitos novos, e consequentemente adquirem maior estabilidade (MOREIRA, 2005).

Na formação de profissionais de saúde, o ensino simulado permite o desenvolvimento de habilidades psicomotoras, de comunicação, a interação entre os domínios cognitivo e afetivo, demonstrando eficácia na educação cognitiva e comportamental (MANO *et al.*, 2019b). Geralmente é utilizada para uma variedade de objetivos de ensino e aprendizagem, como o desenvolvimento de habilidades técnicas, conceitos, competências, resolução de situações problemas, raciocínio clínico, tomada de decisão, trabalho em equipe, resolução de conflitos, entre outros (JEFFRIES;

RIZZOLO, 2006; JEFFRIES; MCNELIS; WHEELER, 2008; HOADLEY, 2009; CENTER, 2013).

Nos processos de avaliação do estudante que passa por práticas simuladas, frequentemente, têm sido destacadas a aquisição de conhecimento, a satisfação, a autoconfiança, a motivação, entre outros aspectos, os quais, usualmente são mensurados através de escalas recomendadas e validadas por diversos autores (BAPTISTA *et al.*, 2014; ALMEIDA *et al.*, 2015). No entanto, a avaliação de forma geral é complexa, ampla e multidimensional. Deve buscar o aprimoramento da execução das habilidades previstas para o cenário simulado e tem como objetivo corrigir deficiências, sanar dificuldades, e manter condições e processos satisfatórios (MANO *et al.*, 2019a; MANO *et al.*, 2019b).

Todavia, muitas vezes, os julgamentos são isolados e possuem relação com fatos, pessoas e fenômenos avaliados. Nesse sentido, a análise das emoções pode expandir o caráter avaliativo (MANO *et al.*, 2019b). Explorar as emoções, pode levar o estudante e o professor a compreender o contexto de forma diferente, com foco nos objetivos definidos para a atividade, além de possibilitar um plano de ação capaz de sustentar o processo de ensino e a aprendizagem decorrente (BRAZEAL; BROWN; COUCH, 2016; MANO *et al.*, 2019a). Nesse contexto, foi realizado esse estudo, que tem como objetivo avaliar as emoções de estudantes de enfermagem durante um cenário clínico simulado de alta fidelidade.

4.5.1 Metodologia na classificação emocional em simulação clínica

O estudo foi desenvolvido durante uma oficina que ocorreu em uma universidade pública brasileira e descreve um estudo quantitativo e descritivo realizado com 24 estudantes de graduação em Enfermagem, regularmente matriculados, de diferentes escolas, com idade superior a 18 anos e envolvidos em um cenário de simulação. O evento foi divulgado online, na página da instituição onde foi realizado. O tema central foi “Cuidados de enfermagem para pacientes clínicos hospitalizados”. A inscrição foi gratuita e todos os alunos receberam material de leitura sobre os temas abordados no evento. Durante a oficina, os alunos foram convidados a participar do estudo e não houve recusas.

Na oficina, os alunos participaram de uma simulação clínica de alta fidelidade usando dramatização (“Cuidados de enfermagem ao paciente com ascite e desconforto respiratório seguido de vômito” - Figura 21). O cenário foi construído, testado e validado, alcançando 100% de consenso de um grupo de especialistas antes de sua aplicação. Os alunos participaram em grupos de três. Após a conclusão do cenário, os alunos foram interrogados e responderam a um questionário com dados pessoais.

Durante o desenvolvimento do cenário, câmeras de vídeo foram colocadas para registrar o comportamento dos alunos para posterior análise e determinação das emoções exibidas. A análise consistiu em três etapas:

Figura 21 – Simulação clínica - “Cuidados de enfermagem ao paciente com ascite e desconforto respiratório seguido de vômito”. Imagem dos estudantes submetidos ao experimento no cenário simulado.

Fonte: Elaborada pelo autor.

1. *Separação*: consistiu na separação do material para análise, isto é, as filmagens foram divididas em imagem, som e texto para aplicação da abordagem multimodal na classificação da emoção;
2. *Extração*: essa etapa foi baseada nas metodologias para extração das características emocionais da face, voz e texto, como descrito nas Seções 4.1, 4.2 e 4.3, respectivamente;
3. *Classificação*: para a classificação foi utilizada a proposta da abordagem multimodal para a classificação da emoção, como descrito na Seção 4.4.

As questões emocionais foram analisadas em seis pontos críticos da simulação clínica, conforme identificado pelos pesquisadores: **1)** comunicação aluno-paciente; **2)** abordagem da queixa do paciente; **3)** avaliação clínica do paciente; **4)** episódio de vômito (distrator atribuído ao cenário); **5)** intervenção de enfermagem (oxigenoterapia) e; **6)** reavaliação do paciente. Como discutido na seção “Conhecimento prévio sobre emoções” (Seção 2.2), essa divisão permitiu que os pesquisadores avaliassem as emoções mais latentes e o tempo gasto em cada ponto crítico do cenário simulado.

Vale salientar, que o estudo foi submetido e aprovado pelo Comitê de Ética em Pesquisa da Escola de Enfermagem de Ribeirão Preto sob a resolução nº 155/2013 (Anexo A). Os sujeitos assinaram o termo de consentimento livre e esclarecido, concedendo aos pesquisadores permissão para usar suas informações pessoais para fins de pesquisa.

4.5.2 Avaliação e discussão

Vinte e quatro estudantes de enfermagem participaram, dos quais 23 eram do sexo feminino (95,8%) e 1 (4,2%) do sexo masculino. A média de idade foi de 22,6 anos, com idade mínima de 19 anos, máxima de 25,4 anos e mediana de 22,9 anos. Entre esses estudantes, 2 estavam no segundo ano do programa, 10 no terceiro ano e 10 no quarto ano. Todos os 24 (100%) já haviam participado de práticas de laboratório e 20 (83,3%) já haviam participado de práticas de simulação.

Os pesquisadores analisaram as emoções identificadas pela abordagem multimodal nos seis pontos críticos previamente identificados na simulação. Ao todo, **83.215 emoções** foram identificadas em todo o cenário de simulação. A Figura 22 ilustra essas observações e a Tabela 12 apresenta os resultados em porcentagens.

Figura 22 – Número de emoções identificadas por momentos na simulação clínica.

Fonte: Elaborada pelo autor.

Tabela 12 – Relação dos momentos da simulação e emoções (%).

Momentos	Emoções						
	Alegria	Desgosto	Medo	Neutro	Raiva	Surpresa	Tristeza
1	15,62	2,51	2,89	19,43	32,90	8,66	17,98
2	5,08	5,10	7,28	20,25	30,29	12,66	19,15
3	11,60	1,95	3,31	45,40	22,82	4,53	10,39
4	6,61	4,66	6,46	8,94	35,62	13,82	23,90
5	4,16	4,31	7,59	13,63	31,16	16,35	22,83
6	6,45	7,22	5,91	10,87	30,69	19,81	19,04

De acordo com os dados da análise, nos seis pontos críticos, os estudantes manifestaram sentimentos diferentes, que foram destacados pelo tempo significativo gasto na atividade no momento 2 (aproximando-se da queixa do paciente) e pelo curto período de tempo gasto no momento 3 (avaliação clínica do paciente), como mostra a Figura 22. As filmagens mostraram que os alunos não esperavam que a queixa da dispneia do paciente se agravasse (momento 2) e estavam muito inseguros quanto à conduta adequada. Na avaliação clínica do paciente (momento 3), realizaram um breve exame físico com foco na cavidade torácica; no entanto, eles logo perceberam que precisavam tomar alguma forma de ação.

Com base na análise das emoções identificadas, a Figura 23 mostra a representação das emoções identificadas no cenário (Table 12) e sua relação com o modelo dimensional das emoções - o Modelo Circumplexo (SCHERER, 2005) (Figura 1).

Figura 23 – Frequência de emoções segundo o Modelo Circumplexo (SCHERER, 2005).

Fonte: Elaborada pelo autor.

No modelo proposto por Scherer (2005), a raiva não é uma emoção restritiva, mas indica uma valência negativa, com alta meta de realização, potencial médio de enfrentamento e alto nível de obstrução do objetivo. A Figura 23 mostrou claramente a hesitação dos alunos sobre o diagnóstico e conduta terapêutica. Os níveis controlados de estresse auxiliam no processo de ensino e aprendizagem e induzem a processos cognitivos precisos (CAHILL; ALKIRE, 2003; MANO *et al.*, 2019b). No entanto, os alunos que não estão bem preparados para o conteúdo teórico ou que não receberam treinamento prático adequado podem apresentar obstrução de objetivos. O estado neutro apresentado no momento 3 (avaliação clínica do paciente) está relacionado a emoções como empatia, paz, paciência e tranquilidade, todas associadas à valência positiva, nível médio-alto de passividade, potencial médio de enfrentamento e alto nível de realização de metas (SCHERER, 2005), que podem ser consideradas altamente adequadas em processos de avaliação clínica, situações mais comuns no cotidiano da prática de enfermagem. No entanto, devido à sua importância na resolução de casos clínicos, é necessário tempo significativo

para a coleta de dados, o que não foi o que ocorreu nesta amostra quando os alunos se depararam com a piora do paciente.

Também vale a pena notar, que menos representações e valores mais baixos de emoções encontradas foram concentrados nos intervalos associados com valência positiva, nível de excitação médio, alto potencial de enfrentamento e alto nível de realização de metas, como felicidade, determinação, interesse, coragem e serenidade. Isso representa maior domínio cognitivo e de atitudes diante do cenário simulado; no entanto, este não foi o caso no presente estudo.

Quando a simulação é usada como método de aprendizagem, técnicas de *debriefing*⁶ são usadas para discutir como os alunos se sentiram durante os cenários, no experimento eles auto-relatam os aspectos positivos e falhas em seu desempenho profissional. Quando as simulações são usadas como uma forma de avaliação, as medidas quantitativas determinam se os alunos passaram ou falharam. No entanto, o reconhecimento de emoções pode ser usado tanto para fins de aprendizado como de avaliação, como parte do *debriefings*. As informações fornecidas por esta tecnologia podem ajudar a identificar pontos de maior dificuldade ou domínio, que podem ser comparados com a análise da autoavaliação dos alunos. Também pode ser usado para avaliar o potencial de enfrentamento nas situações e distrações regulares planejadas em simulações.

4.6 Considerações finais

Segundo os psicólogos, as emoções desempenham um papel crucial no relacionamento e no cotidiano de cada indivíduo. Por isso, acreditamos que esta é uma característica importante a ser medida durante o monitoramento em ambientes residenciais. Como resultado, adotamos uma abordagem multimodal para a classificação emocional, levando em consideração recursos faciais e vocais para validar a ideia em termos de desempenho, precisão e análise estatística. Diante disso, este capítulo investigou o uso da face, tom de voz e descrição da fala para a classificação do estado emocional do usuário.

Mediante a análise facial, exploramos uma metodologia baseada em características geométricas para o reconhecimento do indivíduo e a classificação da emoção. No reconhecimento do indivíduo os experimentos conduzidos obtiveram uma precisão de 99,75% e para a classificação do aspecto emocional a taxa de precisão foi em torno de 76% e houve uma boa convergência com o uso do SVM. Subsequentemente, realizamos a análise do tom de voz, adotando uma metodologia baseada em espectro de energia, para a classificação da emoção. Os experimentos conduzidos demonstraram que, para a classificação do tom de voz, o kNN obteve uma melhor taxa de precisão se comparada com os outros algoritmos testados, em torno de 74%. Por fim, realizamos a análise emocional da descrição da fala, baseada na metodologia *bag of words*, alcançando com a abordagem do N. Bayes uma taxa de precisão de 72% na classificação emocional.

⁶ *Debriefing* é o relatório de um objetivo, projeto ou informações obtidas. É um processo estruturado após o exercício e/ou evento que analisa as ações tomadas.

Visto a discrepância que cada método fornece na classificação emocional e o desafio de classificar as emoções, adotamos uma metodologia baseada na relação multimodal de diferentes componentes. Baseado nas metodologias estudadas para o reconhecimento facial, tom de voz e descrição da fala, adotamos a fusão em nível de características para a classificação da emoção. Os experimentos conduzidos demonstraram que, quando considerada a relação multimodal, há um ganho significativo na taxa de precisão, em torno de 83%, se comparado quando utilizado as abordagens separadamente. Adicionalmente, realizamos um estudo em um ambiente real para a classificação da emoção, demonstrando a utilidade e aplicabilidade do reconhecimento emocional do indivíduo.

Espera-se que esta metodologia para a identificação das emoções possa ser incorporada em uma infraestrutura HSH e forneça, de forma automática e sem a necessidade de intervenção humana, indícios de aspectos emocionais do usuário para o cuidado em ambientes residenciais.

INTELIGÊNCIA COMPUTACIONAL PARA IDENTIFICAÇÃO DE ASPECTOS FÍSICOS

Neste capítulo, é discutido a identificação de aspectos físicos do usuário. Vale salientar, que neste trabalho tratamos aspectos físicos as atividades, com foco em uma possível queda, e a qualidade do ambiente para o repouso, visto que tais aspectos podem afetar, consideravelmente, o cotidiano dos usuários no ambiente residencial. Inicialmente, é apresentada a motivação da identificação das atividades do usuário, a metodologia para a classificação de atividades, bem como a avaliação e discussão dos resultados da classificação. Em seguida, discutimos o monitoramento e a metodologia adotada para a identificação do ambiente para o repouso, e apresentamos os resultados do monitoramento em um ambiente real. Finalmente, apresenta-se as considerações finais deste capítulo.

5.1 Identificação de atividades

O aumento da expectativa de vida resulta no crescimento da população idosa em grandes centros urbanos (MATHERS *et al.*, 2015; MANO *et al.*, 2016b; OLIVEIRA *et al.*, 2018). Embora grande parte da população esteja vivendo mais, ainda é necessário estratégias para auxiliar na saúde e bem-estar de idosos, visto que são mais propensos a acidentes e doenças crônicas (MATHERS *et al.*, 2015; OLIVEIRA *et al.*, 2018). Além disso, apresentam dificuldades ou limitações na realização de tarefas cotidianas, o que pode afetar sua qualidade de vida (MATHERS *et al.*, 2015; OLIVEIRA *et al.*, 2018). Neste cenário, a tecnologia pode ser utilizada com o objetivo de permitir que idosos possam viver de maneira mais independente e segura, por meio de sistemas de monitoramento, auxiliando equipe médica e/ou familiares em situações de riscos (MANO *et al.*, 2016a).

Diversos sistemas têm utilizado o monitoramento dos movimentos, realizados pelos usuários em suas residências, como uma forma de monitorar sua rotina a distância e de modo

não intrusivo (MANO *et al.*, 2016b; NETO *et al.*, 2019). Para que atue de forma não intrusiva, o sistema apenas promove alguma ação quando detecta alterações no padrão de movimentos que um usuário normalmente efetua, portanto, a captura de movimentos precisos e de forma constante é indispensável.

Como exemplo de aplicação de acelerômetro e giroscópio, o serviço *QMedic - Medical Alert System* oferece aos seus usuários uma pulseira com um botão que pode ser acionado sempre que o usuário precisar de assistência médica (como ilustrado na Figura 24). Além de monitorar quedas, o serviço captura os movimentos do usuário de forma passiva antes de executar uma ação (SYSTEM, 2015).

Figura 24 – Exemplo de aplicação de monitoramento de quedas.

Fonte: System (2015).

Esta ação pode ser o envio de informações para um dispositivo móvel, de equipe médica e/ou familiares, tais como mensagens de texto, informando sobre uma situação de risco sofrida pelo indivíduo.

5.1.1 Inteligência Artificial para detecção de quedas

A área de classificação visa construir sistemas que aprendam com os dados. No contexto de monitoramento de sinais, os algoritmos de classificação utilizam o tipo de aprendizado supervisionado. Os algoritmos de aprendizado supervisionado são treinados com exemplos rotulados, isto é, a saída da entrada é conhecida, por exemplo, uma imagem que contém uma face com expressão já conhecida (CLARK; NIBLETT, 1989; MANO *et al.*, 2015). Neste sentido, o acelerômetro e o giroscópio de dispositivos móveis são explorados, de forma conjunta ou separada, frequentemente no contexto de HSH, devido a precisão que podem fornecer nos dados de movimentos capturados e por possuírem um caráter não intrusivo (MANO *et al.*, 2016b; NETO *et al.*, 2019). Assim, pode-se entender suas diferenças da seguinte forma:

- **Acelerômetro:** sensor de movimento que captura a variação na aceleração que um corpo exerce no espaço de forma tridimensional, ou seja, nos eixos X, Y e Z (BONOMI *et al.*, 2009; MANO *et al.*, 2016b). Por meio desses valores, é possível determinar o deslocamento de um corpo em movimento.

- **Giroscópio:** sensor de movimento que captura a mudança de rotação que um corpo exerce no espaço de forma tridimensional, ou seja, nos eixos X, Y e Z (O'DONOVAN; AYER, 2011; MANO *et al.*, 2016b). Por meio desses valores, é possível determinar a rotação e a mudança de direção de um corpo em movimento.

A Figura 25 ilustra a metodologia adotada para classificação das atividades do usuários, considerando a abordagem multimodal para a classificação. Novamente, como descrito na Seção 4.4, para a classificação multimodal, utilizamos a abordagem baseada na fusão em nível de características, isto é, a classificação das atividades é baseada na fusão dos recursos extraídos do acelerômetro e giroscópio do *smartphone* e *smartwatch* do usuário, além do modelo de classificação para combinar as saídas do processamento de características e definir a atividade do usuário, e por consequência a identificação da possível queda do usuário no ambiente residencial.

Figura 25 – Estrutura para a classificação de atividades do usuário, com ênfase em uma possível queda.

Fonte: Elaborada pelo autor.

5.1.2 Avaliação e discussão na detecção de atividades

Considerando o conjunto de dados do acelerômetro e giroscópio de dois dispositivos computacionais distintos (*smartphone* e *smartwatch*) para a identificação das atividades do usuário, foram coletados dados de 3 indivíduos, entre 18 e 70 anos. Essa estratégia foi desenvolvida considerando a necessidade de garantir que ambos os dados, tanto do acelerômetro quanto do giroscópio, fossem adquiridos no mesmo instante de tempo (*timestamp*).

Os experimentos foram realizados de modo não intrusivo para a aquisição dos dados das atividades consideradas nesse estudo (andando, correndo, sentado, deitado, em pé e queda). Os dados coletados correspondem a 15 minutos de cada usuário realizando cada uma das atividades. Por sua vez, os dados foram coletados em um intervalo de tempo correspondente a 20 milissegundos. No final do experimento, foram adquiridos um conjunto de 135.000 dados de cada atividade, **totalizando 810.000 dados**. A metodologia utilizada para os experimentos foi a

realização da média simples de cada 50 linhas de dados, visto que esse número retorna a média do acelerômetro e giroscópio no intervalo de tempo de 1 segundo.

Inicialmente, analisou-se o desempenho dos algoritmos por meio de uma técnica de Planejamento e Avaliação de Experimentos. Nesse caso, foi utilizado novamente o *k-fold cross-validation* com $k = 10$, sendo $k - 1$ para treino e o restante para teste.

Os resultados demonstram que o algoritmo SVM, para esse conjunto de dados, possibilita uma classificação mais precisa, se comparado com os outros classificadores utilizados para comparação. Isso pode ser visto na Figura 26, a qual apresenta os Boxplots referentes as classificações realizadas. É importante ressaltar que o Boxplot com cor vermelha se refere aos resultados do SVM e exibe a mediana da acurácia superior as alcançadas pelos demais classificadores (os valores das medianas podem ser vistas na Tabela 13 e a maior encontra-se destacada). Além disso, é possível notar uma menor dispersão dos resultados obtidos pelo SVM, o que demonstra uma maior estabilidade em suas execuções.

Figura 26 – Boxplots das acurácias apresentadas pelos classificadores para determinar as atividades do usuário.

Fonte: Elaborada pelo autor.

Com objetivo de validar os resultados, foram realizadas análises estatísticas. Inicialmente, utilizamos o método de *Shapiro Wilk* para verificar sua adequação a normalidade e consequentemente conduzir para testes paramétricos ou não-paramétricos. Devido os *p*-valores apresentados na Tabela 13 serem todos maiores que 0.05, consideramos aceita a hipótese de normalidade com confiabilidade de 95%. Portanto, utilizaremos testes paramétricos para as próximas análises.

As comparações de pares realizadas com o teste-*T* são exibidas na Tabela 14. Tais resultados mostram que o SVM não apresenta diferença estatisticamente significativa com os

Tabela 13 – Acurácia média (%) e p -valores (*Shapiro Wilk*) dos algoritmos avaliados.

Classificadores	Acurácia	p -valores (<i>Shapiro Wilk</i>)
Árv. Dec.	93,93	0.8924
Fuzzy	96,33	0.0723
kNN	96,19	0.5251
N. Bayes	94,90	0.6500
SVM	96,62	0.6632

demais classificadores. Contudo, dado a acurácia mediana superior e a menor dispersão dos resultados apresentados pelo SVM, consideramos tal abordagem mais eficiente do que o uso dos demais classificadores para a identificação das atividades do usuário.

Tabela 14 – Comparações pareadas utilizando testes- T na classificação de atividades.

	Árv. Dec.	Fuzzy	kNN	N. Bayes
Fuzzy	1.0	-	-	-
kNN	1.0	1.0	-	-
N. Bayes	1.0	1.0	1.0	-
SVM	0.8	1.0	1.0	1.0

Estes resultados demonstram que, entre os classificadores que foram explorados para esse contexto, o SVM é a técnica que tem uma maior precisão na classificação das atividades de acordo com os sensores utilizados (acelerômetro e giroscópio). A Figura 27 mostra a Matriz de Confusão, que considera a classificação das atividades dos usuários.

Figura 27 – Matriz de Confusão da classificação de atividades (abordagem do SVM).

Fonte: Elaborada pelo autor.

Isso pode fornecer um importante auxílio no cuidado da saúde em ambientes domésticos inteligentes, de forma automática, por meio de uma infraestrutura de IoT. Fizemos uso dos dados combinados do acelerômetro e giroscópio com o intuito de aumentar a precisão na identificação de atividades. Os experimentos foram conduzidos nesta frente e obteve uma precisão de 96.62% ao identificar as atividades (andando, correndo, sentado, deitado, em pé e queda) executadas pelo usuário.

5.2 Qualidade do ambiente para o repouso

No cenário HSH, a utilização de dispositivos com múltiplos sensores pode fornecer, com maior precisão, informações de contexto do usuário, as quais permitem a identificação de situações por meio de processamento contextual, como por exemplo, atividade física, batimento cardíaco, ocorrência de queda, qualidade do sono, entre outros aspectos (PERERA *et al.*, 2013). A ciência de situação representa a capacidade de um sistema computacional obter uma visão abrangente e de alto nível de abstração dos contextos de interesse das aplicações, que pode ser utilizada no processo de tomada de decisão (PERERA *et al.*, 2013).

Segundo Ancoli-Israel *et al.* (1997), algumas características específicas referente ao monitoramento de pessoas idosas estão relacionados as questões de repouso, sendo elas: i.) relatos de idosos que afirmam passar mais tempo na cama devido ao avanço da idade; ii.) apresentam mais despertares durante o período noturno e; iii.) há um aumento significativo de queixa de hipersonia, as quais frequentemente são secundárias a outras doenças. Adicionalmente, idosos tem queixas relacionadas ao sono decorrentes de mudanças fisiológicas específicas do processo de envelhecimento e/ou de doenças que podem causar distúrbios secundários de sono (ANCOLI-ISRAEL, 2005). A experiência de um sono insatisfatório ou insuficiente é bastante desagradável e tem reflexo no desempenho, no comportamento e bem-estar durante as atividades da vida diária (CEOLIM; DIOGO; CINTRA, 2001).

Aliado ao crescimento dos usuários de sistemas computacionais e ao acesso facilitado aos meios de telecomunicação, têm-se a possibilidade de capturar e analisar informações geradas pelo comportamento de indivíduos em diversos contextos (MANO *et al.*, 2016b; MANO *et al.*, 2016a; NETO *et al.*, 2017; KRONBAUER; LUZ; CAMPOS, 2018). A detecção da qualidade do sono pode ser utilizada para o desenvolvimento de novas maneiras de interação do usuário e, posteriormente, caminhar para os dispositivos cognitivos (CAMPBELL; CHOUDHURY, 2012). Tecnologias da informação que possam interagir e auxiliar no cotidiano são um dos caminhos para atender as preferências e as necessidades dos idosos para uma melhor qualidade de vida (MANO *et al.*, 2016b).

5.2.1 Sistema de monitoramento do ambiente - Qualidade do ambiente para o repouso

Pesquisas (YAO *et al.*, 2008; COSTA; CEOLIM, 2013; HORSTEN *et al.*, 2018) nesse sentido buscam o entendimento de fatores que afetam positivamente ou negativamente na qualidade do sono de indivíduos, destacando-se: a iluminação intensa, ocorrência de ruídos e movimentação durante o período noturno. Assim, as informações básicas abordadas neste estudo são: i.) luminosidade; ii.) intensidade do som e; iii.) movimentação.

O sistema desenvolvido permite o monitoramento não-intrusivo através de dispositivos móveis e seus respectivos sensores, permitindo que novos dispositivos sejam integrados para coleta de diferentes informações, a fim de alcançar uma taxa maior de decisões assertivas. A Figura 28 ilustra as camadas lógicas do sistema para o monitoramento do ambiente e as respectivas informações básicas coletadas. Na camada 1 é realizada a aquisição e transmissão dos dados referentes ao ambiente. Posteriormente, na camada 2, é realizado o processamento das características para identificar a qualidade do ambiente para o repouso.

Figura 28 – Contexto não intrusivo para o monitoramento do ambiente para a qualidade do sono.

Fonte: Elaborada pelo autor.

Vale salientar, que o monitoramento não intrusivo realizado pelo sistema se refere ao comportamento transparente ao usuário, o qual não percebe o funcionamento do sistema. Além disso, o sistema apenas realiza alguma atuação (por exemplo, comunica familiares e/ou a equipe médica sobre um longo período de insônia) quando uma situação inadequada é detectada. Situações inadequadas podem ser entendidas como situações que apresentam características específicas ou fora do comportamento padrão do indivíduo, as quais precisam ser comunicadas.

Dado sua natureza, este sistema tem como requisitos a consistência dos dados e a alta disponibilidade do serviço. A consistência do sistema está relacionada ao comportamento estável e com valores reais válidos do ambiente. Já a alta disponibilidade é necessária por se tratar de um sistema de monitoramento e que, caso o sistema se torne indisponível, este não será capaz de identificar situações inadequadas durante esse período.

5.2.2 Metodologia para o monitoramento do ambiente

Como descrito, o funcionamento do sistema de monitoramento do ambiente coleta três informações do ambiente, são eles: i.) luminosidade; ii.) intensidade do som (ruído) do ambiente e; iii.) movimentação do dispositivo. Tais informações podem ser exploradas individualmente ou combinadas entre si devido a facilidade de representação do contexto do ambiente. Nesse sentido, por considerar o conjunto de dados obtidos por quatro sensores distintos (luminosidade, microfone, acelerômetro e giroscópio) com o objetivo de monitorar e classificar as condições de um ambiente, considerando uma noite agradável e confortável para o repouso do idoso, a classificação do ambiente é dada considerando os seguintes rótulos: 1) insatisfatório; 2) moderado e; 3) satisfatório.

Os experimentos apresentados foram avaliados em dois momentos distintos. A primeira avaliação tem como foco a análise da sensibilidade e qualidade dos dados fornecidos pelos sensores, isto é, verificar o quanto cada sensor selecionado é sensível as mudanças do ambiente e se tais mudanças são perceptíveis pelos sensores. Já a segunda avaliação verificou a consistência do sistema no monitoramento por um período contínuo, onde foram estabelecidos quatro cenários distintos, como mostrado no Quadro 7.

Quadro 7 – Cenários do experimento para o monitoramento do ambiente.

Cenário	Luminosidade	Ruído	Movimento
1	Sem presença	Sem presença	Moderado
2	Com iluminação	Sem presença	Moderado
3	Sem presença	Com ruído	Moderado
4	Com iluminação	Com ruído	Moderado

O Quadro 7 representa os ambientes submetidos ao monitoramento do sistema, a saber: 1) ambiente totalmente escuro e sem presença de ruído; 2) ambiente com iluminação e sem presença de ruído; 3) ambiente sem iluminação e com presença de ruído e; 4) ambiente com iluminação e presença de ruído. Ainda, todos os ambientes apresentaram, por um intervalo curto de tempo, algum tipo de movimento, isso para caracterizar acessos ao banheiro, uma noite com grande movimentação ou mesmo insônia.

O ambiente utilizado, para os experimentos, foi um quarto residencial e para variar os aspectos do ambiente foram utilizados: i.) para a iluminação a luz de teto do quarto e a iluminação proporcionada por um dispositivo de vídeo (notebook); ii.) para a geração de ruídos foi utilizado um dispositivo sonoro (tocador de música) que emitia sons constantes ou intercalados; iii.) para o movimento do dispositivo o usuário realizou movimentações intencionais. Ainda, para proporcionar detalhes suficientes visando possíveis replicações dos experimentos, o dispositivo utilizado foi o smartphone Samsung Galaxy S6, Android OS v5.0.2, Octa-core (4x2.1 GHz Cortex-A57 & 4x1.5 GHz Cortex-A53).

5.2.3 Avaliação e discussão dos resultados

Inicialmente, os testes foram conduzidos para verificar a sensibilidade e qualidade dos dados fornecidos pelos sensores. Os resultados apresentados são referentes ao período de 60 minutos de monitoramento, onde é possível verificar a variância dos dados coletados no eixo Y. A Figura 29a apresenta a distribuição dos dados referente a luminosidade do ambiente i.) sem a presença de luz e ii.) com a presença de iluminação gerada pela lâmpada do ambiente. A Figura 29b apresenta a distribuição dos dados referente ao ruído de um ambiente i.) silencioso e ii.) com ruído constante gerado pelo dispositivo de áudio. Por fim, a Figura 29c apresenta a distribuição dos dados referente ao movimento do dispositivo i.) totalmente imóvel e ii.) em movimento realizado pelo usuário. Vale salientar, que para a identificação de movimento faz-se a combinação dos dados obtidos pelo acelerômetro e giroscópio para o cálculo da variância de movimento aplicado.

Figura 29 – Boxplots referente a sensibilidade e qualidade do sensores de luminosidade, ruído e movimento.

Fonte: Elaborada pelo autor.

Estes resultados demonstram que os sensores selecionados apresentam uma sensibilidade aceitável no sentido de identificar as diferenças entre os ambientes. O termo aceitável se refere ao fato dos dados não se sobreporem, possibilitando uma divisão clara entre os dados coletados em cada ambiente. Assim, acredita-se, que estes resultados apresentam indícios de que tais informações permitem monitorar a variação de um ambiente real durante o período noturno.

Com base nestes resultados e nos rótulos para classificar o ambiente durante um determinado momento, possivelmente no período noturno, é definido uma pontuação referente aos intervalos de valores que refletem a característica do ambiente. A Tabela 15 apresenta, de acordo com a sensibilidade verificada na primeira avaliação, a faixa de valores atribuídos para os intervalos de cada característica (**baixo**, **moderado** e **alto**) do ambiente.

Tabela 15 – Relação da faixa de intensidade dos valores dos sensores do monitoramento do ambiente.

Sensores	Faixa	Valor
Luminosidade	0 - 80 (b)	0
	81 - 160 (m)	1
	160 - ... (a)	2
Ruído	0 - 1000 (b)	0
	1001 - 3000 (m)	1
	3001 - ... (a)	2
Movimento	0 - 2 (b)	0
	2.1 - 4 (m)	1
	4.1 - ... (a)	2

As faixas de valores são empregadas na classificação do ambiente, obedecendo os seguintes critérios: resultado da soma dos valores referente a cada característica do ambiente (definidos na coluna Valor da Tabela 15) inferior ou igual a 1 classifica o ambiente como satisfatório para o sono; resultado igual a 2 classifica como ambiente com condições moderadas para o sono e; resultado da soma igual ou superior a 3 classifica o ambiente como insatisfatório para o sono (ver Quadro 8). Por exemplo, um ambiente com luminosidade 1, ruído 0 e movimento 0, é considerado um ambiente satisfatório para o sono, enquanto que um ambiente com luminosidade 1, ruído 0 e movimento 2, é considerado um ambiente insatisfatório para o sono. Para isso, levou-se em consideração as pesquisas (YAO *et al.*, 2008; COSTA; CEOLIM, 2013; HORSTEN *et al.*, 2018) que evidenciam aspectos relacionados a iluminação intensa, geração de ruídos e movimentação durante o período noturno na relação com a qualidade do sono.

Quadro 8 – Classificação do ambiente para o repouso.

Resultado da soma dos valores definidos na Tabela 15	Classificação do ambiente
≤ 1	Satisfatório
$= 2$	Moderado
≥ 3	Insatisfatório

Na segunda fase dos experimentos foi realizada a análise da consistência do sistema ao realizar o monitoramento por um período prolongado, onde foram considerados os cenários apresentados no Quadro 7. A Figura 30 apresenta o período de monitoramento de todos os cenários, os quais tiveram a duração de oito horas, sendo que a cada intervalo de uma hora a classificação foi realizada para verificar a qualidade do ambiente para o repouso. Nestes experimentos o sistema se mostrou consistente e com alta disponibilidade, já que permaneceu

em execução durante todo o período de experimentação e não apresentou valores inadequados ao monitoramento do ambiente.

Figura 30 – Testes do monitoramento dos cenários por um período prolongado.

Fonte: Elaborada pelo autor.

Nota-se, que a partir das avaliações realizadas, tanto para a análise da sensibilidade e qualidade dos dados fornecidos pelos sensores quanto para avaliação da consistência do sistema no monitoramento por um período contínuo, o sistema se mostrou consistente e com alta disponibilidade, conforme seus requisitos definidos anteriormente, podendo fornecer indícios da qualidade do ambiente para o repouso do idoso.

5.3 Considerações finais

Este capítulo investigou o uso de sensores para a identificação do estado físico do usuário, com foco na identificação de atividades e na qualidade do ambiente para o repouso. Na identificação de atividades, exploramos a utilização de dados de sensores para detectar as atividades realizadas por indivíduos no ambiente residencial. Isso pode fornecer um importante auxílio no cuidado da saúde em ambientes domésticos inteligentes, de forma automática, através de uma infraestrutura de IoT (ou seja, sem a necessidade de intervenção humana para detectar novos recursos). Fizemos uso dos dados combinados do acelerômetro e giroscópio com o intuito de aumentar a precisão na identificação das atividades do indivíduo. Os experimentos foram conduzidos nesta frente e obteve uma precisão de 96.62% ao identificar as atividades executadas pelo usuário (andando, correndo, sentado, deitado, em pé e queda).

Por sua vez, para o monitoramento da qualidade do ambiente para o repouso, fez-se o uso de sensores de dispositivos computacionais (luminosidade, ruído e movimento) que

possam monitorar o ambiente para verificar as condições que este se encontra ou mesmo possíveis causas de desconforto. Isso pode fornecer um importante auxílio no cuidado, visto que questões relacionadas ao repouso podem influenciar diretamente o comportamento e bem-estar da população idosa. Os resultados demonstram que o sistema proposto possibilita o monitoramento contínuo e não intrusivo da qualidade do ambiente de maneira coerente a realidade. Além disso, o sistema mostrou ter um funcionamento consistente ao permanecer em execução por longos períodos. Ainda, foi possível verificar a importância da combinação dos dados de variados sensores, fornecendo uma classificação mais assertiva que o uso de um único sensor no que diz respeito ao monitoramento do ambiente para a qualidade do sono.

Espera-se que esta metodologia para a identificação de atividades possa ser incorporada em uma infraestrutura HSH e forneça, de forma automática e sem a necessidade de intervenção humana, indícios do comportamento e do ambiente do usuário para o cuidado em ambientes residenciais.

ARQUITETURA PARA O MONITORAMENTO DE ASPECTOS FÍSICOS E EMOCIONAIS PARA SAÚDE EM CASAS INTELIGENTES

Este capítulo apresenta o *Emotional Level Architecture Communication hEalthcare* (ENLACE) como uma proposta de arquitetura para o monitoramento de aspectos físicos e emocionais no ambiente HSH. Inicialmente, é descrita uma visão geral do ENLACE e a estratégia para seu desenvolvimento. Em seguida, apresentamos o protótipo desenvolvido para o monitoramento residencial, discutimos o monitoramento físico e emocional como um “serviço” e o processo de tomada de decisão de distribuição de carga de processamento do ENLACE. Posteriormente, apresentamos a metodologia adotada para a avaliação de desempenho da arquitetura e os resultados obtidos. Por fim, as considerações finais deste capítulo são apresentadas.

6.1 ENLACE - Arquitetura baseada em camadas para o monitoramento em ambientes residenciais

Visto a importância de fatores físicos e emocionais para melhorar a qualidade de vida das pessoas e permitir que elas vivam de forma mais independente, a aplicação do HSH se mostra uma promissora solução, pois inclui recursos que podem ajudar a identificar e prevenir distúrbios no cotidiano (Mano, 2018). Nesse contexto, um dos aspectos a serem considerados, os serviços projetados para auxiliar o processo de extração de conhecimento devem ser efetivados assim que os dados forem detectados, ou seja, nas bordas da rede. No entanto, os dispositivos integrados ao ambiente HSH, geralmente têm recursos limitados e, portanto, não é possível utilizá-los para determinadas tarefas, sendo necessário utilizar recursos com maior poder de processamento para

explorar os benefícios potenciais de um ambiente residencial inteligente para a saúde (MANO *et al.*, 2016a).

Os paradigmas da computação de borda e de nuvem não devem ser mutuamente exclusivos, mas complementares. Por esse motivo, é essencial avaliar e categorizar os serviços que devem ser fornecidos em cada nível da rede, já que fatores como consumo de energia, volume de dados de transmissão entre camadas e latência de processamento de aplicações devem ser incluídos nesses ambientes (MOHAN; KANGASHARJU, 2016; NETO *et al.*, 2019; MUTLAG *et al.*, 2019; MANO *et al.*, 2019). Diante disso, propomos a arquitetura ENLACE, baseado no uso da IoT para o monitoramento inteligente e individualizado do estado físico e emocional dos indivíduos, onde os sensores são dispositivos no ambiente e/ou de propriedade do usuário, e sua finalidade é coletar dados (por exemplo, imagens e sinais fisiológicos) sobre as condições do cotidiano das pessoas.

Inicialmente, os recursos da borda da rede são definidos para gerenciar e processar informações coletadas. As funções podem ser divididas entre vários tomadores de decisão para fins de escalabilidade, e os recursos de borda servem como uma interface entre o usuário e o cuidador/médico/membros da família. A Figura 31 ilustra a arquitetura proposta para o ENLACE.

Figura 31 – Cenário do ENLACE baseado em três níveis de processamento em tempo de execução.

Fonte: Elaborada pelo autor.

6.1.1 Protótipo para o monitoramento de aspectos físicos e emocionais em ambientes residenciais

Baseado na necessidade de coletar e transmitir dados, foi desenvolvido um dispositivo para monitorar aspectos físicos e emocionais em Casas Inteligentes. A Figura 32 ilustra o equipamento construído, composto por um Arduino pro-Mini e um *Raspberry PI 3* para a coleta,

processamento e transmissão dos dados. O dispositivo apresenta as seguintes características de sensores e atuadores:

- *A* - caixa compacta impressa em material plástico ABS, com dimensões de 99 milímetros de comprimento, 70 milímetros de altura e 43 milímetros de largura;
- *B* - botão de liga/desliga o dispositivo;
- *C* - sensor de presença para o monitoramento do indivíduo;
- *D* - microfone para aquisição da fala do usuário e/ou ruídos do ambiente;
- *E* - câmera infra-vermelho para aquisição de imagens do usuário, tanto no período diurno quanto noturno;
- *F* - *Light Emitting Diode* (LEDs) para informar se o dispositivo está ou não em funcionamento;
- *G* - sensor de temperatura e umidade para o monitoramento do ambiente;
- *H* - sensor de distância para ativação da câmera caso o usuário se aproxime do dispositivo, e;
- *I* - sensor de luminosidade para o monitoramento do ambiente.

Figura 32 – Protótipo do dispositivo de monitoramento residencial.

Fonte: Elaborada pelo autor.

O microcontrolador Arduino pro-Mini é responsável por realizar a interface de comunicação entre sensores (sensor de luminosidade, distância e presença) com o computador. Além disso, o microcontrolador é responsável pelo acionamento da pluralidade de LEDs. O computador Raspberry PI 3 é responsável por receber as informações do microcontrolador e dos sensores

restantes (sensor de câmera, microfone, umidade e temperatura). Além disso, por meio da comunicação Wi-fi, este tem o papel de dispositivo local na arquitetura ENLACE, sendo responsável pela comunicação do dispositivo com o servidor local e a nuvem para o processamento das informações.

6.1.2 Monitoramento físico e emocional como serviço

“*Everything as a Service*” é um modelo que oferece serviços que podem atender às necessidades específicas de clientes ou empresas, onde a principal característica é o serviço prestado e não a tecnologia da informação utilizada (PERERA *et al.*, 2014). O ENLACE foi baseado na abordagem de *Internet of Things Data as a Service Module* (IoTDSM) (BARROS *et al.*, 2018b; MANO *et al.*, 2019). Essa abordagem fornece uma arquitetura de referência para criar um sistema interoperável, permitindo o reconhecimento de aspectos físicos e emocionais como serviço. Este, por sua vez, não depende da camada de aplicação, permitindo executá-lo em ambientes distintos, como dispositivo local, servidor local ou na nuvem. Os serviços avaliados nesse estudo são a análise emocional (abordagem multimodal para classificação da emoção (Seção 4.4)) e a análise de aspectos físicos (identificação de atividade (Seção 5.1) e qualidade do ambiente para o repouso (Seção 5.2)).

A Figura 33 ilustra o fluxo de execução do ENLACE. Primeiramente, os dados são coletados e enviados para o dispositivo local. O dispositivo local executa a tomada de decisão e verifica, com base nos custos, se o processamento será realizado no dispositivo local ou enviado a camadas com maior poder de processamento, ou seja, o servidor local ou a nuvem. Posteriormente, o módulo de classificação, composto pelo serviço de análise física e emocional, processa os dados e envia a saída para o cliente final.

Figura 33 – Fluxo do processo de comunicação entre as camadas do ENLACE para a classificação de aspectos físicos e emocionais do usuário.

Fonte: Elaborada pelo autor.

6.1.3 Tomada de decisão de distribuição de carga entre camadas

No sistema ENLACE, a distribuição da carga de processamento é baseado no impacto dos serviços em suas camadas, isto é, baseado nos custos de processamento. O Algoritmo 1 apresenta o pseudo-código de tomada de decisão para a distribuição da carga de processamento no ENLACE.

Algoritmo 1 – Algoritmo para tomada de decisão de distribuição de carga entre camadas.

Input: novo bloco de dados D (imagens e/ou coordenadas)

```

while  $D_i \in \{D_1, \dots, D_n\}$  do
  if (PrimeiraExecução()) then
     $C_{ld} \leftarrow \text{processingLocalDevice}(D_1)$ 
     $C_{ls} \leftarrow \text{sendLocalServer}(D_1)$ 
     $C_c \leftarrow \text{sendCloud}(D_1)$ 

     $\text{storeLocalDevice}(C_{ld})$ 
     $\text{storeLocalServer}(C_{ls})$ 
     $\text{storeCloudCost}(C_c)$ 
  else
     $eC_{ld} \leftarrow \text{estimateLocalDevice}(D_i)$ 
     $eC_{ls} \leftarrow \text{estimateLocalServer}(D_i)$ 
     $eC_c \leftarrow \text{estimateCloudProcessing}(D_i)$ 

    if ( $eC_{ld} \leq eC_{ls}$ ) then
       $C_{ld} \leftarrow \text{processingLocal}(D_i)$ 
       $\text{storeLocalDevice}(C_{ld})$ 
    else
      if ( $eC_{ls} \leq eC_c$ ) then
         $C_c \leftarrow \text{sendLocalServer}(D_i)$ 
         $\text{storeLocalServer}(C_{ls})$ 
      else
         $C_c \leftarrow \text{sendCloud}(D_i)$ 
         $\text{storeCloudCost}(C_c)$ 
      end
    end
  end
end
end

```

Na primeira execução, ou seja, na primeira vez que o sistema inicia a operação, os serviços são executados em todos os ambientes (dispositivo local, servidor local e nuvem) para aferição dos custos de processamento (consumo de energia e o tempo de resposta na execução) da aplicação. A Equação 6.1 é usada para estimar o custo de execução das aplicações:

$$C_x = \int_1^x V_x I_x t_x + \text{Cost}_{T_x + R_x} \quad (6.1)$$

onde C_x é a taxa de transmissão da interface de rede, V_x é a voltagem do dispositivo e I_x é a corrente elétrica durante um intervalo de tempo t_x (MANO *et al.*, 2019). Deve-se observar que, a execução do serviço no servidor local e na nuvem, deve considerar o tempo de ida e volta nas transmissões entre os dispositivos. Assim, $Cost$ é uma constante que representa $(T_x + R_x)$, onde T_x é o custo de transmissão e R_x é o custo de recebimento. Ambos, T_x e R_x , são definidos pela voltagem do dispositivo e pela corrente elétrica (NETO *et al.*, 2019).

A cada nova execução, o ENLACE compara os custos da primeira execução e os valores estimados da nova amostra de dados, realizando a decisão de descarregamento em relação à camada de processamento. O algoritmo equilibra o processamento nas camadas do ENLACE para fornecer um melhor desempenho, pois o tempo de execução pode variar dependendo da carga de trabalho. Vale ressaltar, que o algoritmo de tomada de decisão é executado na camada de dispositivo local, e este decide se os dados serão enviados para as próximas camadas ou se serão processados na camada atual.

6.1.4 Avaliação e discussão

Esta seção descreve a metodologia e os resultados utilizados para avaliar os serviços no ambiente de HSH. Tais serviços estão sujeitos as variações na carga de trabalho e podem ser executados individualmente ou juntos, em qualquer camada da arquitetura.

6.1.4.1 Configuração do experimento

Para avaliar o desempenho do ENLACE, o Quadro 9 apresenta as informações referente a capacidade computacional dos dispositivos da arquitetura nos diferentes níveis (dispositivo local, servidor local e ambiente de nuvem).

Quadro 9 – Configurações do ambiente do ENLACE.

Ambiente	Processador	Memória
Dispositivo local	4 x ARM Cortex-A53, 1.2GHz	1GB LPDDR2
Servidor local	Intel Core 2 Quad 9400, 2.66GHz	8 GB DIMM
Cloud	MD FX 8-Core, 4.0Ghz	32 GB RAM DDR3

Devido à discrepância do poder computacional que pode ser encontrado em cada camada da arquitetura ENLACE, diferentes cargas de trabalho foram aplicadas a todas elas. Ao gerar a carga de trabalho, números diferentes de requisições de serviço são realizados em um instante de tempo, seguindo uma distribuição de probabilidade uniforme para representar o monitoramento contínuo da arquitetura. O Quadro 10 apresenta o número de requisições atribuídos no experimento em cada nível do ambiente.

Quadro 10 – Configurações do experimento - taxa de requisições.

	Dispositivo local	Servidor local	Cloud
Total de requisições	200	200	200
Requisições por minuto	16, 24, 32	32, 64, 96	64, 96, 128

6.1.4.2 Resultados experimentais e discussão

Os resultados se concentram na análise de processamento e tempo de resposta das taxas de requisições nos diferentes níveis da arquitetura ENLACE. Embora o monitoramento do usuário seja simulado, o ambiente de processamento do ENLACE foi implementado como esperado de um ambiente real. Os resultados mostram os testes realizados para cada serviço individual, bem como para a execução dos serviços de forma conjunta. As Figuras 34a, 34b e 34c mostram as representações gráficas do tempo de resposta para executar cada taxa de requisição dos serviços no nível do dispositivo local.

Figura 34 – Número de solicitações e tempo decorrido dos serviços no nível do dispositivo local.

(c) Serviços de monitoramento físico e emocional.

Fonte: Elaborada pelo autor.

Em geral, o tempo de resposta é diretamente proporcional à taxa de requisição, o que significa que quanto mais pedidos são realizados, o tempo de resposta se torna mais significativo. Com relação aos valores das taxas de solicitações, a taxa de carga de trabalho de 16 e 24 solicitações por minuto, têm praticamente o mesmo tempo de resposta. Por outro lado, o tempo de resposta para uma carga de trabalho de 32 solicitações por minuto aumenta em um tempo linear, com proporção maior à medida que mais solicitações são feitas por minuto, como pode ser visualizado na Figura 34c. Conjuntamente, o tempo de resposta das últimas solicitações diminui em uma taxa linear, pois toda a carga de trabalho (200 solicitações) é executada e o *host* manipula somente as solicitações na fila.

É essencial determinar como cada serviço influenciou o tempo de resposta. A Figura 34a mostra a representação gráfica do tempo de resposta necessário para cada requisição executar o monitoramento físico. Este gráfico mostra que, independentemente da carga de trabalho e da taxa de solicitação por minuto, o tempo de resposta é baixo e constante; isso pode ser explicado pelas características dos algoritmos que não requerem processamento intensivo. Por outro lado, a representação gráfica do tempo de resposta para a realização do monitoramento emocional, conforme mostrado na Figura 34b, mostra que a taxa de carga de trabalho de 16 e 24 solicitações por minuto, o tempo de resposta é constante, o que sugere que o dispositivo pode manipular mais solicitações do que o número de solicitações recebidas. Quando a taxa de cobrança aumenta para 32 solicitações por minuto, o tempo de resposta aumenta em um tempo linear, o que sugere que o número de solicitações excede sua capacidade de processamento, o que significa que essas solicitações devem aguardar para serem processadas.

Posteriormente, as Figuras 35a, 35b e 35c mostram as representações gráficas do tempo de resposta necessário para executar cada taxa de requisição dos serviços no nível do servidor local.

O comportamento dos resultados apresentados no gráfico da Figura 35c é similar com os resultados da Figura 34c, pois podem operar ambos os serviços, embora utilize diferentes níveis de taxas de requisições. Ao contrário do tempo de resposta no dispositivo local, onde há uma taxa de carga de trabalho de 32 solicitações por minuto e o tempo de resposta cresce em uma escala linear, ocorre um tempo constante no servidor local em relação a 32 e 64 solicitações por minuto. Isso sugere que, devido ao maior poder computacional, o tempo de resposta para atender às solicitações é menor se comparado ao cenário anterior (dispositivo local). Por outro lado, uma taxa de carga de trabalho de 96 solicitações, ocasiona uma sobrecarga no servidor local e o tempo de resposta aumenta em uma escala linear. As Figuras 35a e 35b mostram os tempos de resposta da taxas de requisições ao realizar o processamento dos serviços separadamente. Como no cenário anterior, o mesmo comportamento foi observado no nível do dispositivo local, onde, para o serviço de monitoramento físico, o tempo de resposta é baixo e constante, enquanto que para o serviço de monitoramento emocional requer mais recursos computacionais, causando sobrecarga e aumentando o tempo de resposta.

Figura 35 – Número de solicitações e tempo decorrido dos serviços no nível do servidor local.

(a) Serviço de monitoramento físico.

(b) Serviço de monitoramento emocional.

(c) Serviços de monitoramento físico e emocional.

Fonte: Elaborada pelo autor.

Finalmente, as Figuras 36a, 36b e 36c mostram as representações gráficas do tempo de resposta para executar cada taxa de requisição dos serviços no nível do servidor na nuvem.

Como pode ser visto na Figura 36c, o poder computacional do servidor na nuvem, permite um tempo baixo constante para uma taxa de carga de trabalho de 64 e 96 solicitações por minuto, diferente do servidor local que tinha uma taxa de crescimento linear para 96 solicitações por minuto. No entanto, para uma taxa de carga de trabalho de 128 solicitações por minuto, o tempo de resposta aumenta em escala linear, como no dispositivo local e o servidor local, quando submetido a uma carga de trabalho alta. De maneira semelhante ao cenário anterior, o tempo de resposta individual de cada serviço mostra que, para o serviço de monitoramento físico, ocorre um tempo de resposta constante e baixo (Figura 36a), enquanto o serviço de monitoramento emocional requer mais recursos computacionais, pois causam sobrecarga no servidor de nuvem e aumentam o tempo de resposta (Figura 36b).

Figura 36 – Número de solicitações e tempo decorrido dos serviços no nível do servidor em nuvem.

(a) Serviço de monitoramento físico.

(b) Serviço de monitoramento emocional.

(c) Serviços de monitoramento físico e emocional.

Fonte: Elaborada pelo autor.

Em resumo, os resultados mostram que o tempo de resposta está diretamente relacionado ao poder computacional, às características dos algoritmos e ao número de solicitações por minuto submetidos a cada camada da arquitetura. Como mostrado na análise, o algoritmo de monitoramento físico requer pouco poder computacional e, independentemente do número de solicitações por minuto definido em cada nível do ENLACE, não é possível sobrecarregar o dispositivo alvo. Por outro lado, o algoritmo de monitoramento emocional, que requer mais recursos computacionais, comporta-se de maneira diferente, dependendo do número de solicitações por minuto. Para este serviço, quanto maior o número de solicitações por minuto, maior a probabilidade de sobrecarregar o dispositivo de destino. Além disso, o mesmo número de solicitações por minuto pode ter um tempo de resposta diferente para cada nível do ENLACE, dependendo de ocorrer ou não uma sobrecarga. Acima de tudo, deve-se notar que os tempos de resposta linear não são desejáveis no ambiente HSH, uma vez que o tempo de resposta atrasado podem levar a uma tomada de decisão errada.

É essencial entender o comportamento dos algoritmos, em cada camada do ENLACE, para o processo de tomada de decisão. Os resultados obtidos em cada nível mostram que, quando há uma determinada taxa de requisições de serviço, é aconselhável executá-los em um nível com maior poder de processamento, ou até mesmo, dividir os pedidos entre os diferentes níveis das camadas da arquitetura, obtendo um nível mais aceitável de tempo de resposta para monitoramento físico e emocional em ambientes de HSH.

6.2 Considerações finais

Inúmeros são os métodos, técnicas e instrumentos que apoiam no monitoramento para a avaliação da saúde do usuário, isso devido à oportunidade de sistemas computacionais reagirem ao estado e a mudança do comportamento do ambiente ou do próprio indivíduo, permitindo que aplicações computacionais se tornem ciente do contexto do usuário. Diante disso, este capítulo investigou a proposta de uma arquitetura HSH para a identificação do estado físico e emocional do usuário.

O ENLACE foi proposto como uma arquitetura baseada em camadas para processar dados de sensores (por exemplo, características faciais e atividades físicas) para detectar e classificar o estado físico e emocional do usuário no ambiente HSH. Os resultados demonstram, que a arquitetura baseada em níveis de processamento, permite uma melhor distribuição de carga de serviço, em vista da complexidade de cada algoritmo e da capacidade de processamento de cada camada da arquitetura. Os resultados mostram que o algoritmo para aspectos físicos mantém um tempo de resposta constante e baixo, enquanto o algoritmo para a identificação de aspectos emocionais é mais suscetível ao crescimento linear, dependendo da taxa de solicitações em cada camada da arquitetura. Assim, com relação a carga de requisições dos serviços, pode-se estabelecer qual a camada da arquitetura será mais adequada para realizar o processamento, uma vez que pode fornecer uma resposta em tempo hábil para qualquer tipo de intervenção.

CONCLUSÃO

Os sistemas para Saúde em Casas Inteligentes permitem o uso de tecnologias para o monitoramento e, consequentemente, o cuidado para com a saúde de pessoas com alguma deficiência, mobilidade reduzida e idosos. Vale ressaltar, que o cuidado da saúde não envolve somente questões relacionadas com aspectos físicos, mas também com aspectos emocionais. O fator emocional tem ganhado destaque, tanto no setor acadêmico/científico quanto médico e social, visto o aumento de doenças e/ou enfermidades causadas por questões emocionais. Nesse contexto, na área de Computação, inúmeros são os métodos, técnicas, instrumentos e estratégias que apoiam a avaliação do comportamento, devido à oportunidade de sistemas computacionais reagirem ao estado físico e emocional de um indivíduo.

Devido as lacunas identificadas na literatura em relação a integração e desenvolvimento de modelos para o reconhecimento de aspectos físicos e emocionais, além de ambientes residenciais inteligentes para o monitoramento e processamento de informações relacionadas a saúde, esta Tese apresentou soluções para os desafios encontrados, por meio de metodologias para o tratamento e classificação de aspectos físicos e emocionais do usuário e por meio de uma arquitetura dinâmica, baseada em camadas de processamento, para ser aplicada em ambientes para HSH.

As avaliações utilizaram métodos estatísticos paramétricos e não-paramétricos para analisar os resultados obtidos. Os resultados mostraram alcançar o objetivo desta Tese, sendo cinco deles relevantes: **i.)** metodologias para extração de características da expressão facial, tom de voz e descrição da fala para classificação da emoção; **ii.)** metodologias para a classificação de atividades do usuário e identificação da qualidade do ambiente para o repouso; **iii.)** modelo baseado na abordagem multimodal, com a utilização de sensores IoT, na classificação de aspectos físicos e emocionais do indivíduo; **iv.)** desenvolvimento do sistema de monitoramento e dispositivo de baixo custo para saúde em casas inteligentes, e; **v.)** arquitetura (ENLACE) de comunicação, baseada em camadas de processamento, para ambientes HSH. Portanto, por meio

dos resultados apresentados, pode-se fornecer um sistema residencial eficiente e inteligente, que pode prover o monitoramento de aspectos físicos e emocionais, e melhorar a interação, qualidade de vida e bem-estar de pessoas com deficiências, mobilidade reduzida e idosos.

7.1 Aplicabilidade e trabalhos futuros

O modelo multimodal para o reconhecimento de aspectos físicos e emocionais, juntamente com a arquitetura multi-camadas para o processamento, pode ser utilizado/aplicado em diversos contextos para o auxílio do usuário, além do ambiente residencial demonstrado neste estudo, pode auxiliar em ambientes mais complexos, como em ambientes hospitalares e/ou casas de repouso, para fornecer indícios de doenças causadas por emoções, falta de expressões emocionais ou instabilidade de emoções, como esquizofrenia, depressão, autismo e transtorno bipolar.

Além disso, no ambiente educacional, o modelo multimodal de reconhecimento de emoções pode melhorar a experiência dos alunos em um ambiente de aprendizado, ajudar os professores em metodologias de ensino, bem como programas de treinamento virtual, como mostrado nos experimentos descritos na Seção 4.5.

Apesar das inúmeras aplicações e bons resultados dos modelos de classificação de aspectos físicos e emocionais do usuário e da comunicação e processamento da arquitetura ENLACE, sua precisão e desempenho podem ser melhoradas. Portanto, pretendemos explorar:

- outras abordagens que identificam emoções, por exemplo, Tendência Comportamental - redes sociais e Reações Fisiológicas - batimento cardíaco;
- outros níveis de fusão, como por exemplo a fusão em nível decisão e/ou fusão híbrida;
- *middleware* para melhorar a comunicação e processamento na arquitetura ENLACE;
- arquitetura *fog computing* e *micro-fog*, e;
- estratégias de interação/persuasão que melhor atendam às preferências dos usuários.

7.2 Principais contribuições

As principais contribuições desta Tese foram apresentadas em forma de documentos científicos que foram publicados em conferências e periódicos nacionais e internacionais. A seguir são apresentadas as contribuições e publicações provenientes desta Tese:

- Publicações em periódicos:

1. **Mano, L. Y.**; Barros, V. A.; Nunes, L. H.; Sawada, L. O.; Estrella, J. C.; Ueyama, J. (2019). *ENLACE: A Combination of Layer-Based Architecture and Wireless Communication for Emotion Monitoring in Healthcare*. Mobile Information Systems. DOI: <https://doi.org/10.1155/2019/7329187>.
2. **Mano, L. Y.**; Mazzo, A.; Neto, J. R. T.; et al. (2019). *Using emotion recognition to assess simulation-based learning*. Nurse Education in Practice. DOI: <https://doi.org/10.1016/j.nepr.2019.02.017>.
3. **Mano, L. Y.**; Mazzo, A.; Neto, J. R. T.; et al. (2019). *The relation of satisfaction, self-confidence and emotion in a simulated environment*. International Journal of Nursing Education Scholarship. DOI: <https://doi.org/10.1515/ijnes-2018-0009>.
4. Neto, J. R. T.; Filho, G. P. R.; **Mano, L. Y.**; Villas, L. A. and Ueyama, J. (2019). *Exploiting offloading in IoT-based micro-fog: experiments with face recognition and fall detection*. Wireless Communications and Mobile Computing. DOI: <https://doi.org/10.1155/2019/2786837>.
5. Sawada, L. O.; **Mano, L. Y.**; Neto, J. R. T. and Ueyama, J. (2019). *A module-based framework to emotion recognition by speech: a case study in clinical simulation*. Journal of Ambient Intelligence and Humanized Computing. DOI: <https://doi.org/10.1007/s12652-019-01280-8>.
6. Neto, J. R. T.; Filho, G. P. R.; **Mano, L. Y.**; Ueyama, J. (2018). *VERBO: Voice Emotion Recognition dataBase in Portuguese Language*. Journal of Computing Science. DOI: <https://doi.org/10.3844/jcssp.2018.1420.1430>.
7. Filho, G. P. R.; **Mano, L. Y.**; Valejo, A.; Villas, L. A.; Ueyama, J. (2018). *A Low-Cost Smart Home Automation to Enhance Decision-Making based on Fog Computing and Computational Intelligence*. IEEE Latin America Transactions. DOI: <https://doi.org/10.1109/TLA.2018.8291472>.
8. Ueyama, J.; Façal, B. S.; **Mano, L. Y.**; Bayer, G.; et al. (2017). *Enhancing Reliability in Wireless Sensor Networks for Adaptive River Monitoring Systems: Reflections on our Long-term Deployment in Brazil*. Computers, Environment and Urban Systems. DOI: <https://doi.org/10.1016/j.compenvurbsys.2017.05.001>.
9. Façal, B. S.; Freitas, H.; Gomes, P.; **Mano, L.**; et al. (2017). *An adaptive approach for UAV-based pesticide spraying in dynamic environments*. Computers and Electronics in Agriculture. DOI: <https://doi.org/10.1016/j.compag.2017.04.011>.
10. **Mano, L. Y.**; Vasconcelos, E.; Ueyama J. (2016). *Identifying Emotions in Speech Patterns: Adopted Approach and Obtained Results*. Latin America Transactions. DOI: <https://doi.org/10.1109/TLA.2016.7817010>.

11. **Mano, L. Y.**, Sawada, L. O. and Ueyama J. (2016). *Abordagem para a interação afetiva: um estudo de caso com player de música*. Journal on Advances in Theoretical and Applied Informatics.
DOI: <https://doi.org/10.26729/jadi.v2i2.2108>.
 12. **Mano, L. Y.**; Funes, M. M.; Volpato, T. and Neto, J. R. T. (2016). *Explorando tecnologias de IoT no contexto de Health Smart Home: uma abordagem para detecção de quedas em pessoas idosas*. Journal on Advances in Theoretical and Applied Informatics.
DOI: <https://doi.org/10.26729/jadi.v2i1.1667>.
- Publicações em conferências e simpósios
 1. **Mano, L. Y.** (2018). *Emotional condition in the Health Smart Homes environment: emotion recognition using ensemble of classifiers*. IEEE International Conference on INnovations in Intelligent SysTems and Applications, INISTA, Grécia.
 2. Toneti, B. F., **Mano, L. Y.**, Sawada, L. O; Oliveira, I. G.; et al. (2018). *Desenvolvimento de tecnologias digitais voltadas ao ensino de uma prática integrativa em saúde*. Congresso Internacional de Tecnologia e Humanização na Comunicação em Saúde, Brasil¹.
 3. Junior, C. R. G.; Gomes, P. H.; **Mano, L. Y.**; Oliveira, R. B.; et al. (2017). *A Machine Learning-Based Approach for Prediction of Plant Protection Product Deposition*. 2017 Brazilian Conference on Intelligent Systems, BRACIS, Brazil.
 4. **Mano, L. Y.**, Sawada, L. O; Franzon, J. C.; et al. (2017). *Abordagem baseada na análise do comportamento emocional para avaliação clínica*. 3º Congresso de Graduação da USP, Brasil.
 5. Filho, C. K. C.; Mazzo A.; Machado, G. C. C.; Franzon, J. C.; Meska, M. H. G.; **Mano, L. Y.**; et al. (2017). *Empatia de estudantes de enfermagem nas práticas clínicas simuladas: assistência de enfermagem nas necessidades de eliminações*. 3º Congresso de Graduação da USP, Brasil.
 6. Mazzo A.; Meska, M. H. G.; Franzon, J. C.; **Mano, L. Y.**; et al. (2017). *Construção e validação de cenários clínicos para a formação de profissionais de saúde*. 3º Congresso de Graduação da USP, Brasil.
 7. Bernardi, F. A.; Santos, M. A.; Ferreira, W. D.; Senju, A. L. A.; Jesus, T. H. P.; **Mano, L. Y.**; et al. (2017). *Checklist informatizado em dispositivos móveis para avaliação e feedback imediato em cenários de simulação clínica*. 3º Congresso de Graduação da USP, Brasil.

¹ Trabalho premiado com Menção Honrosa no evento.

8. Júnior, G. A. P.; Mazzo, A.; Santos, M. A.; Senju, A. L. A.; Bernardi, F. A.; Ferreira, W. D.; Jesus, T. H. P.; **Mano, L. Y.**; et al. (2017). *Aprendizagem baseada em projetos no Ensino interprofissional da assistência e educação permanente da rede de atenção de urgência*. 3º Congresso de Graduação da USP, Brasil.
9. Neto, J. R. T.; Filho, G. P. R.; **Mano, L. Y.**; Ueyama, J. (2017). *INCA: Um sistema healthcare flexível baseado no paradigma fog computing e publish/subscribe*. Workshop de Computação Urbana (CoUrb) - 35º Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos, 35º SBRC, Brazil.
10. Sawada, L. O.; **Mano, L. Y.**; Simões, E. V. (2017). *Abordagem baseada em sistemas embarcados para capacitação de profissionais de saúde*. 1º Simpósio Internacional de Simulação Clínica Interprofissional, Brasil.
11. Meska, M. H. G.; Mazzo, A.; Franzon, J. C.; Machado, G. C. C.; **Mano, L. Y.**; et al. (2017). *Construção e validação de cenário com presença de odores*. 1º Simpósio Internacional de Simulação Clínica Interprofissional, Brasil.
12. Bonafim, C. M.; **Mano, L. Y.**; Mazzo, A. (2017). *Análise do diálogo para avaliação do comportamento emocional em cenários simulados*. 1º Simpósio Internacional de Simulação Clínica Interprofissional, Brasil.
13. Nakamura, L. H. V.; Façal, B.; **Mano, L. Y.**; Ueyama, J. (2016). *MFBL - Monitoring Feelings to Better Lives*. VI Brazilian Symposium on Computing Systems Engineering, SBESC, Brazil.

- Produção técnica

1. **Patente**: Ueyama, J.; Vieira, H. F.; Neto, J. R. T.; **Mano, L. Y.** “*Sistema de Monitoramento, Dispositivo de Aquisição e Transmissão de Dados e Método de Identificação e Classificação de Aspectos Físicos e Emocionais*”. Número do registro: BR102018016161-0, Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial, 2018.
2. **Registro de software**: Ueyama, J.; Façal, B. S.; Bayer, G.; **Mano, L. Y.** “*E-Noe - Software de uma Rede de Sensores sem fio para o Monitoramento de rios urbanos*”. Número do registro: BR512016001777-5, data de registro: 22/12/2016, Instituição de registro: INPI - Instituto Nacional da Propriedade Industrial, 2016.

- Revisão de artigos

1. 2019 International Conference on Artificial Intelligence, Information Processing and Cloud Computing - AIIPCC2019, 2019.
2. 11th Latin-American Conference on Communications - LATINCOM, 2019.
3. Simpósio Brasileiro de Games e Entretenimento Digital - SBGames, 2019.

4. 46º Seminário Integrado de Software e Hardware - SEMISH, 2019.
5. 11º Simpósio Brasileiro de Computação Ubíqua e Pervasiva - SBCUP, 2019.
6. 24th IEEE Symposium on Computers and Communications, 2019.
7. IEEE Access, 2018.
8. Journal Heliyon, 2018.
9. IEEE Internet of Things Journal, 2018.
10. 2nd International Conference on Computer Science and Application Engineering - CSAE, 2018.
11. 10º Simpósio Brasileiro de Computação Ubíqua e Pervasiva - SBCUP, 2018.
12. IETE (Institution of Electronics and Telecommunication Engineers) Journal of Research, 2017.
13. Journal on Advances in Theoretical and Applied Informatics, 2017.
14. 9º Simpósio Brasileiro de Computação Ubíqua e Pervasiva - SBCUP, 2017.
15. 1º Simpósio Internacional de Simulação Clínica Interprofissional, 2017.
16. XXXV Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos - SBRC, 2017.
17. 1º Etapa Simpósio Internacional de Iniciação Científica da USP - SIICUSP, 2016.

REFERÊNCIAS

- AARTS, E.; GROTENHUIS, F. Ambient intelligence 2.0: Towards synergetic prosperity. **Ambient Intelligence**, Springer, p. 1–13, 2009. Citado na página 30.
- ABOWD, G. D.; BOBICK, A. F.; ESSA, I. A.; MYNATT, E. D.; ROGERS, W. A. The aware home: A living laboratory for technologies for successful aging. In: **Proceedings of the AAAI-02 Workshop “Automation as Caregiver**. [S.l.: s.n.], 2002. p. 1–7. Citado nas páginas 46 e 48.
- AGUIAR, E. J. de; FAIÇAL, B. S.; UEYAMA, J.; SILVA, G. C.; MENOLLI, A. Análise de sentimento em redes sociais para a língua portuguesa utilizando algoritmos de classificação. In: SBC. **Anais do XXXVI Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos**. [S.l.], 2018. Citado na página 64.
- ALAM, F.; RICCARDI, G. Predicting personality traits using multimodal information. In: **ACM. Proceedings of the 2014 ACM multi media on workshop on computational personality recognition**. [S.l.], 2014. p. 15–18. Citado na página 36.
- ALAM, M. R.; REAZ, M. B. I.; ALI, M. A. M. A review of smart homes—past, present, and future. **IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)**, IEEE, v. 42, n. 6, p. 1190–1203, 2012. Citado na página 45.
- ALMEIDA, R. G. d. S.; MAZZO, A.; MARTINS, J. C. A.; BAPTISTA, R. C. N.; GIRÃO, F. B.; MENDES, I. A. C. Validation to portuguese of the scale of student satisfaction and self-confidence in learning. **Revista latino-americana de enfermagem**, SciELO Brasil, v. 23, n. 6, p. 1007–1013, 2015. Citado na página 73.
- ALVES, L. Y. M. **FlexPersuade-Explorando uma abordagem flexível em softwares de persuasão: um estudo de caso com players de música**. Tese (Doutorado) — Universidade de São Paulo, 2016. Citado nas páginas 27, 31, 32, 33, 34, 35, 39, 40, 42, 48, 49, 51, 52, 53 e 54.
- AMAN, S.; SZPAKOWICZ, S. Identifying expressions of emotion in text. In: **SPRINGER. International Conference on Text, Speech and Dialogue**. [S.l.], 2007. p. 196–205. Citado na página 64.
- ANCOLI-ISRAEL, S. Sleep and aging: prevalence of disturbed sleep and treatment considerations in older adults. **The Journal of clinical psychiatry**, v. 66, p. 24–30, 2005. Citado nas páginas 45 e 84.
- ANCOLI-ISRAEL, S.; POCETA, J. S.; STEPNOWSKY, C.; MARTIN, J.; GEHRMAN, P. Identification and treatment of sleep problems in the elderly. **Sleep medicine reviews**, Elsevier, v. 1, n. 1, p. 3–17, 1997. Citado na página 84.
- AUGUSTO, J.; KRAMER, D.; ALEGRE, U.; COVACI, A.; SANTOKHEE, A. The user-centred intelligent environments development process as a guide to co-create smart technology for people with special needs. **Univers. Access Inf. Soc.**, Springer-Verlag, Berlin, Heidelberg, v. 17, n. 1, p. 115–130, mar. 2018. ISSN 1615-5289. Citado nas páginas 47 e 48.

- AUGUSTO, J. C.; MCCULLAGH, P.; MCCLELLAND, V.; WALKDEN, J. Enhanced healthcare provision through assisted decision-making in a smart home environment. In: **2nd Workshop on Artificial Intelligence Techniques for Ambient Intelligence**. [S.l.: s.n.], 2007. Citado nas páginas 25, 26, 27, 43 e 45.
- BAIENSON, J. N.; PONTIKAKIS, E. D.; MAUSS, I. B.; GROSS, J. J.; JABON, M. E.; HUTCHERSON, C. A.; NASS, C.; JOHN, O. Real-time classification of evoked emotions using facial feature tracking and physiological responses. **International journal of human-computer studies**, Elsevier, v. 66, n. 5, p. 303–317, 2008. Citado na página 35.
- BAPTISTA, R. C. N.; MARTINS, J. C. A.; PEREIRA, M. F. C. R.; MAZZO, A. Students' satisfaction with simulated clinical experiences: validation of an assessment scale. **Revista latino-americana de enfermagem**, SciELO Brasil, v. 22, n. 5, p. 709–715, 2014. Citado na página 73.
- BARROS, P.; CHURAMANI, N.; LAKOMKIN, E.; SIQUEIRA, H.; SUTHERLAND, A.; WERMTER, S. The omg-emotion behavior dataset. **arXiv preprint arXiv:1803.05434**, 2018. Citado na página 67.
- BARROS, V. A.; ESTRELLA, J. C.; PRATES, L. B.; BRUSCHI, S. M. An iot-daas approach for the interoperability of heterogeneous sensor data sources. In: **ACM. Proceedings of the 21st ACM International Conference on Modeling, Analysis and Simulation of Wireless and Mobile Systems**. [S.l.], 2018. p. 275–279. Citado na página 94.
- BONOMI, A. G.; PLASQUI, G.; GORIS, A. H.; WESTERTERP, K. R. Improving assessment of daily energy expenditure by identifying types of physical activity with a single accelerometer. **Journal of Applied Physiology**, Am Physiological Soc, v. 107, n. 3, p. 655–661, 2009. Citado na página 80.
- BRAZEAL, K. R.; BROWN, T. L.; COUCH, B. A. Characterizing student perceptions of and buy-in toward common formative assessment techniques. **CBE-Life Sciences Education**, Am Soc Cell Biol, v. 15, n. 4, p. ar73, 2016. Citado na página 73.
- CADDICK, Z. A.; GREGORY, K.; ARSINTESCU, L.; FLYNN-EVANS, E. E. A review of the environmental parameters necessary for an optimal sleep environment. **Building and environment**, Elsevier, v. 132, p. 11–20, 2018. Citado na página 42.
- CAHILL, L.; ALKIRE, M. T. Epinephrine enhancement of human memory consolidation: interaction with arousal at encoding. **Neurobiology of learning and memory**, Elsevier, v. 79, n. 2, p. 194–198, 2003. Citado na página 76.
- CAI, G.; XIA, B. Convolutional neural networks for multimedia sentiment analysis. In: **Natural Language Processing and Chinese Computing**. [S.l.]: Springer, 2015. p. 159–167. Citado na página 36.
- CAMPBELL, A.; CHOUDHURY, T. From smart to cognitive phones. **Pervasive Computing, IEEE**, v. 11, n. 3, 2012. Citado na página 84.
- CARVALHO, S. T.; ERTHAL, M.; MARELI, D.; SZTAJNBERG, A.; COPETTI, A.; LOQUES, O.; JANEIRO-RJ-BRASIL, R. de. Monitoramento remoto de pacientes em ambiente domiciliar. **XXVIII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos-Salao de Ferramentas**, Gramado, RS, Brasil, p. 1005–1012, 2010. Citado nas páginas 27, 44 e 45.

CASTILLO, J. C.; CASTRO-GONZÁLEZ, Á.; FERNÁNDEZ-CABALLERO, A.; LATORRE, J. M.; PASTOR, J. M.; FERNÁNDEZ-SOTOS, A.; SALICHS, M. A. Software architecture for smart emotion recognition and regulation of the ageing adult. **Cognitive Computation**, Springer, v. 8, n. 2, p. 357–367, 2016. Citado nas páginas 46 e 48.

CENTER, S. I. R. **Simulation Innovation Resource Center**. 2013. <<http://www.sirc.nln.org/mod/glossary/view.php>>. Accessed in November de 2013. Disponível em: <<http://www.sirc.nln.org/mod/glossary/view.php>>. Citado nas páginas 72 e 73.

CEOLIM, M. F.; DIOGO, M. J. D.; CINTRA, F. A. Qualidade do sono de pessoas idosas atendidas no grupo de atenção à saúde do idoso do hospital das clínicas da universidade estadual de campinas. **Nursing (São Paulo)**, v. 4, n. 33, p. 25–9, 2001. Citado nas páginas 45 e 84.

CHAN, M.; ESTÈVE, D.; ESCRIBA, C.; CAMPO, E. A review of smart homes—present state and future challenges. **Computer methods and programs in biomedicine**, Elsevier, v. 91, n. 1, p. 55–81, 2008. Citado na página 45.

CHANEL, G.; KIERKELS, J. J.; SOLEYMANI, M.; PUN, T. Short-term emotion assessment in a recall paradigm. **International Journal of Human-Computer Studies**, Elsevier, v. 67, n. 8, p. 607–627, 2009. Citado na página 35.

CHUNG, W.-Y.; OH, S.-J. Remote monitoring system with wireless sensors module for room environment. **Sensors and Actuators B: Chemical**, Elsevier, v. 113, n. 1, p. 64–70, 2006. Citado nas páginas 44 e 45.

CLARK, P.; NIBLETT, T. The cn2 induction algorithm. **Machine learning**, Springer, v. 3, n. 4, p. 261–283, 1989. Citado nas páginas 34, 35, 36 e 80.

CORCHADO, J. M.; BAJO, J.; TAPIA, D. I.; ABRAHAM, A. Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare. **IEEE Transactions on Information Technology in Biomedicine**, v. 14, n. 2, p. 234–240, March 2010. ISSN 1089-7771. Citado nas páginas 43 e 45.

COSTA, S. V. d.; CEOLIM, M. F. Factors that affect inpatients' quality of sleep. **Revista da Escola de Enfermagem da USP**, SciELO Brasil, v. 47, n. 1, p. 46–52, 2013. Citado nas páginas 85 e 88.

COSTANTINI, G.; IADEROLA, I.; PAOLONI, A.; TODISCO, M. Emovo corpus: an italian emotional speech database. In: EUROPEAN LANGUAGE RESOURCES ASSOCIATION (ELRA). **International Conference on Language Resources and Evaluation (LREC 2014)**. [S.l.], 2014. p. 3501–3504. Citado na página 60.

CRISTIANINI, N.; SHAW-TAYLOR, J. **An introduction to support vector machines and other kernel-based learning methods**. [S.l.]: Cambridge university press, 2000. Citado nas páginas 34 e 35.

CRUZ-ALBARRAN, I. A.; BENITEZ-RANGEL, J. P.; OSORNIO-RIOS, R. A.; MORALES-HERNANDEZ, L. A. Human emotions detection based on a smart-thermal system of thermographic images. **Infrared Physics & Technology**, Elsevier, v. 81, p. 250–261, 2017. Citado nas páginas 40 e 42.

DESMET, P. A multilayered model of product emotions. **The Design Journal**, Bloomsbury Journals (formerly Berg Journals), v. 6, n. 2, p. 4–13, 2003. Citado nas páginas 34 e 51.

- D'MELLO, S. K.; KORY, J. A review and meta-analysis of multimodal affect detection systems. **ACM Computing Surveys (CSUR)**, ACM, v. 47, n. 3, p. 43, 2015. Citado na página 36.
- DOHR, A.; MODRE-OPSRIAN, R.; DROBICS, M.; HAYN, D.; SCHREIER, G. The internet of things for ambient assisted living. In: IEEE. **Information Technology: New Generations (ITNG), 2010 Seventh International Conference on**. [S.l.], 2010. p. 804–809. Citado nas páginas 25 e 30.
- DUDA, P. E.; RICHARD, O. **Hart, Pattern Classification and Scene Analysis**. [S.l.]: John Wiley and Sons, New York, 1973. Citado nas páginas 34 e 35.
- EBNER, N. C.; RIEDIGER, M.; LINDENBERGER, U. Faces—a database of facial expressions in young, middle-aged, and older women and men: Development and validation. **Behavior research methods**, 2010. Citado na página 55.
- EKMAN, P. **Cross-cultural studies of facial expression**. [S.l.]: New York: Academic, 1973. Citado nas páginas 15, 32, 33, 39, 42 e 52.
- _____. **Darwin and facial expression: A century of research in review**. [S.l.]: Ishk, 2006. Citado nas páginas 15, 32, 33, 34 e 64.
- EKMAN, P.; FRIESEN, W. V. The repertoire of nonverbal behavior: Categories, origins, usage, and coding. **Nonverbal communication, interaction, and gesture**, p. 57–106, 1981. Citado na página 31.
- EKMAN, P.; FRIESEN, W. V.; HAGER, J. C. Facs investigator's guide. **A human face**, p. 96, 2002. Citado nas páginas 40 e 42.
- EKMAN, P.; KELTNER, D. Universal facial expressions of emotion. **Seegerstrale U, P. Molnar P, eds. Nonverbal communication: Where nature meets culture**, p. 27–46, 1997. Citado nas páginas 39 e 42.
- EKMAN, P.; ROSENBERG, E.; HAGER, J. **Facial action coding system affect interpretation dictionary (FACSAID)**. 1998. Citado nas páginas 40 e 42.
- EKMAN, R. **What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)**. [S.l.]: Oxford University Press, USA, 1997. Citado nas páginas 40 e 42.
- ELLSWORTH, P. C.; SCHERER, K. R. Appraisal processes in emotion. **Handbook of affective sciences**, v. 572, p. V595, 2003. Citado na página 31.
- EYBEN, F.; WENINGER, F.; GROSS, F.; SCHULLER, B. Recent developments in opensmile, the munich open-source multimedia feature extractor. In: ACM. **Proceedings of the 21st ACM international conference on Multimedia**. [S.l.], 2013. p. 835–838. Citado na página 59.
- FACELI, K. **Inteligência artificial: uma abordagem de aprendizado de máquina**. [S.l.]: Grupo Gen-LTC, 2011. Citado na página 35.
- FAST, L. A.; FUNDER, D. C. Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. **Journal of personality and social psychology**, American Psychological Association, v. 94, n. 2, p. 334, 2008. Citado na página 63.

FERNÁNDEZ-CABALLERO, A.; LATORRE, J. M.; PASTOR, J. M.; FERNÁNDEZ-SOTOS, A. Improvement of the elderly quality of life and care through smart emotion regulation. In: SPRINGER. **International Workshop on Ambient Assisted Living**. [S.l.], 2014. p. 348–355. Citado nas páginas 46 e 48.

FILHO, P. G.; UEYAMA, J.; VILLAS, L. A.; PINTO, A.; GONÇALVES, V. P.; SERAPHINI, S. An intelligent approach for improving energy efficiently in smart grids. In: THE STEERING COMMITTEE OF THE WORLD CONGRESS IN COMPUTER SCIENCE, COMPUTER ENGINEERING AND APPLIED COMPUTING (WORLDCOMP). **Proceedings of the International Conference on Wireless Networks (ICWN)**. [S.l.], 2013. p. 1. Citado na página 34.

FIORI, M.; VESELY-MAILLEFER, A. K. Emotional intelligence as an ability: theory, challenges, and new directions. In: **Emotional Intelligence in Education**. [S.l.]: Springer, 2018. p. 23–47. Citado nas páginas 36 e 66.

FONTAINE, J. R.; POORTINGA, Y. H.; SETIADI, B.; MARKAM, S. S. Cognitive structure of emotion terms in indonesia and the netherlands. **Cognition & Emotion**, Taylor & Francis, v. 16, n. 1, p. 61–86, 2002. Citado nas páginas 31, 33 e 34.

FREDRICKSON, B. L.; LEVENSON, R. W. Positive emotions speed recovery from the cardiovascular sequelae of negative emotions. **Cognition and Emotion**, v. 12, n. 2, p. 191–220, 1998. Citado na página 26.

FRIESEN, E.; EKMAN, P. Facial action coding system: a technique for the measurement of facial movement. **Palo Alto**, Consulting Psychologists Press, v. 3, 1978. Citado nas páginas 39 e 42.

FRIJDA, N. H. The emotions: Studies in emotion and social interaction. **Paris: Maison de Sciences de l'Homme**, 1986. Citado na página 31.

FURUI, S. **Digital speech processing: synthesis, and recognition**. [S.l.]: CRC Press, 2018. Citado na página 59.

GALTERIO, M.; SHAVIT, S.; HAYAJNEH, T. A review of facial biometrics security for smart devices. **Computers**, Multidisciplinary Digital Publishing Institute, v. 7, n. 3, p. 37, 2018. Citado na página 52.

GONÇALVES, V. P.; COSTA, E. P.; VALEJO, A.; FILHO, P. G.; JOHNSON, T. M.; PESSIN, G.; UEYAMA, J. Enhancing intelligence in multimodal emotion assessments. **Applied Intelligence**, Springer, v. 46, n. 2, p. 470–486, 2017. Citado nas páginas 27, 36, 39, 41, 42, 66 e 67.

GONÇALVES, V. P.; GIANCRISTOFARO, G. T.; FILHO, P. G.; JOHNSON, T.; CARVALHO, V.; PESSIN, G.; NERIS, V. P. de A.; UEYAMA, J. Assessing users' emotion at interaction time: a multimodal approach with multiple sensors. **Soft Computing**, Springer, v. 21, n. 18, p. 5309–5323, 2017. Citado nas páginas 41 e 42.

GONÇALVES, V. P.; NERIS, V.; MORANDINI, M.; NAKAGAWA, E. Y.; UEYAMA, J. Uma revisão sistemática sobre métodos de avaliação de usabilidade aplicados em software de telefones celulares. In: BRAZILIAN COMPUTER SOCIETY. **Proceedings of the 10th Brazilian Symposium on on Human Factors in Computing Systems and the 5th Latin American Conference on Human-Computer Interaction**. [S.l.], 2011. p. 197–201. Citado na página 39.

- GONÇALVES, V. P.; NERIS, V.; UEYAMA, J.; GIANCRISTOFARO, G. T.; SERAPHINI, S. *et al.* Emoções na interação humano-computador: um estudo considerando sensores. In: BRAZILIAN COMPUTER SOCIETY. **Proceedings of the 12th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.], 2013. p. 252–255. Citado nas páginas 26 e 35.
- GONÇALVES, V. P.; NERIS, V. P.; UEYAMA, J.; SERAPHINI, S.; DIAS, T. C.; FILHO, G. R. Senior citizens in interaction with mobile phones: A flexible middleware approach to support the diversity. **Proc. of the SERP 2013**, v. 1, p. 56–62, 2013. Citado na página 25.
- GONÇALVES, V. P.; UEYAMA, J. Um estudo sobre o design, a implementação e a avaliação de interfaces flexíveis para idosos em telefones celulares. 2012. Citado nas páginas 27 e 52.
- GOOGLE. **Google Home**. 2018. <https://store.google.com/product/google_home>. Disponível em: <https://store.google.com/product/google_home>. Citado na página 26.
- GUBBI, J.; BUYYA, R.; MARUSIC, S.; PALANISWAMI, M. Internet of things (iot): A vision, architectural elements, and future directions. **Future Generation Computer Systems**, Elsevier, v. 29, n. 7, p. 1645–1660, 2013. Citado na página 30.
- GUR, R. E.; KOHLER, C. G.; RAGLAND, J. D.; SIEGEL, S. J.; LESKO, K.; BILKER, W. B.; GUR, R. C. Flat affect in schizophrenia: relation to emotion processing and neurocognitive measures. **Schizophrenia bulletin**, MPRC, v. 32, n. 2, p. 279–287, 2006. Citado na página 26.
- HELAL, S.; MANN, W.; KING, J.; KADDOURA, Y.; JANSEN, E. *et al.* The gator tech smart house: A programmable pervasive space. **Computer**, IEEE, v. 38, n. 3, p. 50–60, 2005. Citado nas páginas 46 e 48.
- HIRSH, J. B.; PETERSON, J. B. Personality and language use in self-narratives. **Journal of research in personality**, Elsevier, v. 43, n. 3, p. 524–527, 2009. Citado na página 63.
- HOADLEY, T. A. Learning advanced cardiac life support: A comparison study of the effects of low-and high-fidelity simulation. **Nursing education perspectives**, LWW, v. 30, n. 2, p. 91–95, 2009. Citado na página 73.
- HORSTEN, S.; REINKE, L.; ABSALOM, A.; TULLEKEN, J. Systematic review of the effects of intensive-care-unit noise on sleep of healthy subjects and the critically ill. **British journal of anaesthesia**, Elsevier, v. 120, n. 3, p. 443–452, 2018. Citado nas páginas 85 e 88.
- IBM. **Speech to Text, About Speech to Text, IBM Watson Developer Cloud**. 2018. <<https://www.ibm.com/watson/developercloud/doc/speech-to-text/index.html>>. Disponível em: <<https://www.ibm.com/watson/developercloud/doc/speech-to-text/index.html>>. Citado na página 63.
- IBM, I. B. M. **International Business Machines - IBM**. 2018. <<https://www.ibm.com>>. Disponível em: <<https://www.ibm.com>>. Citado na página 26.
- II, N. **Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions**. [S.l.]: Bryssel, 2014. Citado na página 31.
- INTEL. **Internet-of-Things (IoT) Advances Home Healthcare for Seniors**. 2016. <http://www.embeddedintel.com/special_features.php?article=2721>. Acessado em Junho de 2016. Disponível em: <http://www.embeddedintel.com/special_features.php?article=2721>. Citado na página 26.

IVANOVA, A. Humor recognition in psychiatric patients and artificial intelligence. In: **2012 AAAI Fall Symposium Series**. [S.l.: s.n.], 2012. Citado na página 26.

JAIN, N.; KUMAR, S.; KUMAR, A.; SHAMSOLMOALI, P.; ZAREAPOOR, M. Hybrid deep neural networks for face emotion recognition. **Pattern Recognition Letters**, Elsevier, 2018. Citado nas páginas 40 e 42.

JEFFRIES, P. R.; MCNELIS, A. M.; WHEELER, C. A. Simulation as a vehicle for enhancing collaborative practice models. **Critical Care Nursing Clinics of North America**, Elsevier, v. 20, n. 4, p. 471–480, 2008. Citado na página 73.

JEFFRIES, P. R.; RIZZOLO, M. A. Designing and implementing models for the innovative use of simulation to teach nursing care of ill adults and children: A national, multi-site, multi-method study [summary report]. National League for Nursing, 2006. Citado na página 73.

KANG, B.; KIM, D.; CHOO, H. Internet of everything: A large-scale autonomic iot gateway. **IEEE Transactions on Multi-Scale Computing Systems**, v. 3, n. 3, p. 206–214, July 2017. ISSN 2332-7766. Citado nas páginas 47 e 48.

KATSI, C. D.; KATERTSIDIS, N.; GANIATSAS, G.; FOTIADIS, D. I. Toward emotion recognition in car-racing drivers: A biosignal processing approach. **Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on**, IEEE, v. 38, n. 3, p. 502–512, 2008. Citado na página 35.

KHANCHANDANI, K.; HUSSAIN, M. A. Emotion recognition using multilayer perceptron and generalized feed forward neural network. CSIR, 2009. Citado nas páginas 27, 40 e 42.

KRONBAUER, A. H.; LUZ, H. C. da; CAMPOS, J. Mobile security monitor: A wearable computing platform to detect and notify falls. **IEEE Latin America Transactions**, IEEE, v. 16, n. 3, p. 957–965, 2018. Citado na página 84.

KÜHNER, D. "internet der dinge telekommunikationsinfrastruktur. **Seminarband: Mobile und Verteilte Systeme-Ubiquitous Computing Teil IV. Seminarband: Mobile und Verteilte Systeme-Ubiquitous Computing Teil IV. Universität Karlsruhe-Fakultät für Informatik**, p. 1–15, 2007. Citado na página 29.

KUNZE, C.; HOLTMANN, C.; SCHMIDT, A.; STORK, W. Kontextsensitive technologien und intelligente sensorik für ambient-assisted-living-anwendungen. **Ambient Assisted Living-AAL**, VDE VERLAG GmbH, 2008. Citado na página 30.

LANGNER, O.; DOTTSCH, R.; BIJLSTRA, G. *et al.* Presentation and validation of the radboud faces database. **Cognition and Emotion**, 2010. Citado na página 55.

LAPIERRE, N.; NEUBAUER, N.; MIGUEL-CRUZ, A.; RINCON, A. R.; LIU, L.; ROUSSEAU, J. The state of knowledge on technologies and their use for fall detection: A scoping review. **International journal of medical informatics**, Elsevier, v. 111, p. 58–71, 2018. Citado na página 42.

LEE, C. H.; KIM, K.; SEO, Y. S.; CHUNG, C. K. The relations between personality and language use. **The Journal of general psychology**, Taylor & Francis, v. 134, n. 4, p. 405–413, 2007. Citado na página 63.

- LIKAMWA, R.; LIU, Y.; LANE, N. D.; ZHONG, L. Moodscope: Building a mood sensor from smartphone usage patterns. In: ACM. **Proceeding of the 11th annual international conference on Mobile systems, applications, and services**. [S.l.], 2013. p. 389–402. Citado nas páginas 40 e 42.
- LITTLEWORT, G.; WHITEHILL, J.; WU, T.; FASEL, I.; FRANK, M.; MOVELLAN, J.; BARTLETT, M. The computer expression recognition toolbox (cert). In: IEEE. **Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on**. [S.l.], 2011. p. 298–305. Citado nas páginas 35 e 36.
- LOOS, H. M. V. D.; ULLRICH, N.; KOBAYASHI, H. Development of sensate and robotic bed technologies for vital signs monitoring and sleep quality improvement. **Autonomous Robots**, Springer, v. 15, n. 1, p. 67–79, 2003. Citado nas páginas 44 e 45.
- LUCEY, P.; COHN, J. F. *et al.* The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In: **Computer Vision and Pattern Recognition Workshops, IEEE Computer Society Conf. on**. [S.l.: s.n.], 2010. Citado nas páginas 53 e 55.
- MAHLKE, S.; MINGE, M. Consideration of multiple components of emotions in human-technology interaction. In: **Affect and emotion in human-computer interaction**. [S.l.]: Springer, 2008. p. 51–62. Citado nas páginas 33, 34, 37, 51 e 66.
- MAIA, P.; BAFFA, A.; CAVALCANTE, E.; DELICATO, F. C.; BATISTA, T.; PIRES, P. F. A middleware platform for integrating devices and developing applications in e-health. In: IEEE. **Computer Networks and Distributed Systems (SBRC), 2015 XXXIII Brazilian Symposium on**. [S.l.], 2015. p. 10–18. Citado nas páginas 35, 43 e 45.
- MANDAL, M. K.; PANDEY, R.; PRASAD, A. B. Facial expressions of emotions and schizophrenia: A review. **Schizophrenia bulletin**, National Institute of Mental Health, v. 24, n. 3, p. 399, 1998. Citado na página 26.
- Mano, L. Y. Emotional condition in the health smart homes environment: emotion recognition using ensemble of classifiers. In: **2018 Innovations in Intelligent Systems and Applications (INISTA)**. [S.l.: s.n.], 2018. p. 1–8. Citado nas páginas 25, 27, 40, 46, 48, 49, 52, 53, 54 e 91.
- MANO, L. Y.; BARROS, V. A.; NUNES, L. H.; SAWADA, L. O.; ESTRELLA, J. C.; UEYAMA, J. Enlace: A combination of layer-based architecture and wireless communication for emotion monitoring in healthcare. **Mobile Information Systems**, Hindawi, v. 2019, 2019. Citado nas páginas 48, 49, 92, 94 e 96.
- MANO, L. Y.; FAIÇAL, B. S.; NAKAMURA, L. H.; GOMES, P. H.; LIBRALON, G. L.; MENEQUETE, R. I.; FILHO, G. P.; GIANCRISTOFARO, G. T.; PESSIN, G.; KRISHNAMACHARI, B.; UEYAMA, J. Exploiting iot technologies for enhancing health smart homes through patient identification and emotion recognition. **Computer Communications**, v. 89-90, p. 178 – 190, 2016. ISSN 0140-3664. Internet of Things Research challenges and Solutions. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0140366416300688>>. Citado nas páginas 25, 26, 27, 30, 34, 42, 44, 45, 46, 48, 49, 79, 84 e 92.
- MANO, L. Y.; FUNES, M.; VOLPATO, T.; NETO, J. Explorando tecnologias de iot no contexto de health smart home: uma abordagem para detecção de quedas em pessoas idosas. **Journal on Advances in Theoretical and Applied Informatics**, v. 2, n. 1, p. 46–57, 2016. ISSN 2447-5033. Disponível em: <<https://www.revista.univem.edu.br/jadi/article/view/1667>>. Citado nas páginas 25, 26, 27, 30, 35, 42, 43, 45, 48, 49, 79, 80, 81 e 84.

- MANO, L. Y.; GIANCRISTOFARO, G. T.; FAICAL, B. S.; LIBRALON, G. L.; PESSIN, G.; GOMES, P. H.; UEYAMA, J. Exploiting the use of ensemble classifiers to enhance the precision of user's emotion classification. In: **Proceedings of the 16th International Conference on Engineering Applications of Neural Networks (INNS)**. New York, NY, USA: ACM, 2015. (EANN '15), p. 5:1–5:7. ISBN 978-1-4503-3580-5. Citado nas páginas 34, 40, 42, 48, 49, 52, 53 e 80.
- MANO, L. Y.; MAZZO, A.; NETO, J. R.; F., C. K. C.; G., V. P.; UEYAMA, J.; JÚNIOR, G. A. The relation of satisfaction, self-confidence and emotion in a simulated environment. **International Journal of Nursing Education Scholarship**, v. 16, n. 1, 2019. Citado nas páginas 33, 34, 48, 49 e 73.
- MANO, L. Y.; MAZZO, A.; NETO, J. R.; MESKA, M. H.; GIANCRISTOFARO, G. T.; UEYAMA, J.; JÚNIOR, G. A. Using emotion recognition to assess simulation-based learning. **Nurse Education in Practice**, v. 36, p. 13 – 19, 2019. ISSN 1471-5953. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1471595317301324>>. Citado nas páginas 32, 33, 34, 48, 49, 72, 73 e 76.
- MANO, L. Y.; SAWADA, L.; UEYAMA, J. Abordagem para a interação afetiva: um estudo de caso com player de música. **Journal on Advances in Theoretical and Applied Informatics**, v. 2, n. 2, p. 47–54, 2016. ISSN 2447-5033. Disponível em: <<https://www.revista.univem.edu.br/jadi/article/view/2108>>. Citado nas páginas 48, 49 e 52.
- MANO, L. Y.; VASCONCELOS, E.; UEYAMA, J. Identifying emotions in speech patterns: Adopted approach and obtained results. **IEEE Latin America Transactions**, v. 14, n. 12, p. 4775–4780, Dec 2016. ISSN 1548-0992. Citado nas páginas 48, 49, 58, 59 e 60.
- MARTIN, J. H.; JURAFSKY, D. **Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition**. [S.l.]: Pearson/Prentice Hall Upper Saddle River, 2009. Citado na página 59.
- MATHERS, C. D.; STEVENS, G. A.; BOERMA, T.; WHITE, R. A.; TOBIAS, M. I. Causes of international increases in older age life expectancy. **The Lancet**, Elsevier, v. 385, n. 9967, p. 540–548, 2015. Citado na página 79.
- MEHL, M. R.; GOSLING, S. D.; PENNEBAKER, J. W. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. **Journal of personality and social psychology**, American Psychological Association, v. 90, n. 5, p. 862, 2006. Citado na página 63.
- MENA-CHALCO, J.; CESAR-JR, R.; VELHO, L. Banco de dados de faces 3d: Impa-face3d. **IMPA-RJ, Tech. Rep**, 2008. Citado na página 55.
- METALLINO, A.; WOLLMER, M.; KATSAMANIS, A.; EYBEN, F.; SCHULLER, B.; NARAYANAN, S. Context-sensitive learning for enhanced audiovisual emotion classification. **IEEE Transactions on Affective Computing**, IEEE, v. 3, n. 2, p. 184–198, 2012. Citado na página 37.
- MOHAMED, A. G.; DILIP, G.; DEVI, N. R.; ANUSUYA, S. Intelligent healthcare system using internet of things. **International Journal of Pure and Applied Mathematics**, v. 115, n. 8, p. 347–352, 2017. Citado nas páginas 46 e 48.

- MOHAN, N.; KANGASHARJU, J. Edge-fog cloud: A distributed cloud for internet of things computations. In: **2016 Cloudification of the Internet of Things (CIoT)**. [S.l.: s.n.], 2016. p. 1–6. Citado na página 92.
- MOREIRA, M. A. Aprendizaje significativo crítico. **Indivisa: Boletín de estudios e investigación**, Escuela Universitaria La Salle, n. 6, p. 83–102, 2005. Citado na página 72.
- MUTLAG, A. A.; GHANI, M. K. A.; ARUNKUMAR, N.; MOHAMMED, M. A.; MOHD, O. Enabling technologies for fog computing in healthcare iot systems. **Future Generation Computer Systems**, v. 90, p. 62 – 78, 2019. ISSN 0167-739X. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0167739X18314006>>. Citado na página 92.
- NERIS, V. P. de A.; BARANAUSKAS, M. C. C. Making interactive systems more flexible: an approach based on users' participation and norms. In: BRAZILIAN COMPUTER SOCIETY. **Proceedings of the IX Symposium on Human Factors in Computing Systems**. [S.l.], 2010. p. 101–110. Citado na página 27.
- NETO, J. R. T.; FILHO, G. P. R.; MANO, L. Y.; UEYAMA, J. Inca: Um sistema healthcare flexível baseado no paradigma fog computing e publish/subscribe. In: **Anais do I Workshop de Computação Urbana (COURB 2017)**. Porto Alegre, RS, Brasil: SBC, 2017. v. 1. ISSN 2595-2706. Disponível em: <<http://ojs.sbc.org.br/index.php/courb/article/view/2572>>. Citado nas páginas 27, 43, 45 e 84.
- NETO, J. R. T.; FILHO, G. P. R.; MANO, L. Y.; UEYAMA, J. Verbo: voice emotion recognition database in portuguese language. **Journal of Computer Science**, Science Publications, v. 14, n. 11, p. 1420–1430, 2018. Citado nas páginas 40, 42, 48, 49, 59 e 60.
- NETO, J. R. T.; FILHO, G. P. R.; MANO, L. Y.; VILLAS, L. A.; UEYAMA, J. Exploiting offloading in iot-based microfog: Experiments with face recognition and fall detection. **Wireless Communications and Mobile Computing**, Hindawi, v. 2019, 2019. Citado nas páginas 35, 47, 48, 49, 80, 92 e 96.
- NIEDENTHAL, P.; KRAUTH-GRUBER, S.; RIC, F. Psychology of emotion. **Psychology of Emotion**, Psychology Press New York, 2006. Citado na página 33.
- O'DONOVAN, K.; AYER, S. Real-time joint angle measurement using the shimmer wireless sensor platform. In: ACM. **Proceedings of the First ACM Workshop on Mobile Systems, Applications, and Services for Healthcare**. [S.l.], 2011. p. 7. Citado na página 81.
- ØHRN, A.; ROWLAND, T. Rough sets: a knowledge discovery technique for multifactorial medical outcomes. **American journal of physical medicine & rehabilitation**, LWW, v. 79, n. 1, p. 100–108, 2000. Citado na página 35.
- OLIVEIRA, R. N.; ROTH, V.; HENZEN, A. F.; SIMAO, J. M.; NOHAMA, P.; WILLE, E. C. G. Notification oriented paradigm applied to ambient assisted living tool. **IEEE Latin America Transactions**, IEEE, v. 16, n. 2, p. 647–653, 2018. Citado nas páginas 35, 45 e 79.
- OMATU, S.; ROCHA, M. P.; BRAVO, J.; FERNÁNDEZ, F.; CORCHADO, E.; BUSTILLO, A.; CORCHADO, J. M. Distributed computing, artificial intelligence, bioinformatics, soft computing, and ambient assisted living. **Lecture Notes in Computer Science**, v. 5518, 2009. Citado nas páginas 27, 35, 43 e 45.

OSTIR, G. V.; BERGES, I.-M.; OTTENBACHER, M. E.; CLOW, A.; OTTENBACHER, K. J. Associations between positive emotion and recovery of functional status following stroke. **Psychosomatic medicine**, v. 70, n. 4, p. 404–409, 2008. Citado na página 26.

O'TOOLE, A. J.; CASTILLO, C. D.; PARDE, C. J.; HILL, M. Q.; CHELLAPPA, R. Face space representations in deep convolutional neural networks. **Trends in cognitive sciences**, Elsevier, 2018. Citado nas páginas 40 e 42.

PARKINSON, B. **Ideas and realities of emotion**. [S.l.]: Psychology Press, 1995. Citado na página 31.

PATEL, P.; ALI, M. I.; SHETH, A. On using the intelligent edge for iot analytics. **IEEE Intelligent Systems**, v. 32, n. 5, p. 64–69, September 2017. ISSN 1541-1672. Citado nas páginas 47 e 48.

PEÑA-LÓPEZ, I. *et al.* Itu internet report 2005: the internet of things. ITU, 2005. Citado na página 30.

PENNEBAKER, J. W.; KING, L. A. Linguistic styles: language use as an individual difference. **Journal of personality and social psychology**, American Psychological Association, v. 77, n. 6, p. 1296, 1999. Citado na página 63.

Pereira Rocha Filho, G.; Yukio Mano, L.; Demetrius Baria Valejo, A.; Aparecido Villas, L.; Ueyama, J. A low-cost smart home automation to enhance decision-making based on fog computing and computational intelligence. **IEEE Latin America Transactions**, v. 16, n. 1, p. 186–191, Jan 2018. ISSN 1548-0992. Citado nas páginas 48 e 49.

PERERA, C.; ZASLAVSKY, A.; CHRISTEN, P.; GEORGAKOPOULOS, D. Context aware computing for the internet of things: A survey. **IEEE communications surveys & tutorials**, IEEE, v. 16, n. 1, p. 414–454, 2013. Citado na página 84.

_____. Sensing as a service model for smart cities supported by internet of things. **Transactions on Emerging Telecommunications Technologies**, Wiley Online Library, v. 25, n. 1, p. 81–93, 2014. Citado na página 94.

PETER, C.; URBAN, B. Emotion in human-computer interaction. In: **Expanding the Frontiers of Visual Analytics and Visualization**. [S.l.]: Springer, 2012. p. 239–262. Citado nas páginas 27, 35, 41 e 42.

PONS, G.; MASIP, D. Supervised committee of convolutional neural networks in automated facial expression analysis. **IEEE Transactions on Affective Computing**, IEEE, v. 9, n. 3, p. 343–350, 2018. Citado nas páginas 40 e 42.

PORIA, S.; CAMBRIA, E.; BAJPAI, R.; HUSSAIN, A. A review of affective computing: From unimodal analysis to multimodal fusion. **Information Fusion**, Elsevier, v. 37, p. 98–125, 2017. Citado nas páginas 27, 36, 39, 41, 42, 66 e 67.

PORIA, S.; CAMBRIA, E.; GELBUKH, A. Deep convolutional neural network textual features and multiple kernel learning for utterance-level multimodal sentiment analysis. In: **Proceedings of the 2015 conference on empirical methods in natural language processing**. [S.l.: s.n.], 2015. p. 2539–2544. Citado na página 36.

- PORIA, S.; CAMBRIA, E.; HOWARD, N.; HUANG, G.-B.; HUSSAIN, A. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. **Neurocomputing**, Elsevier, v. 174, p. 50–59, 2016. Citado nas páginas 27, 36, 41 e 42.
- PORIA, S.; CAMBRIA, E.; HUSSAIN, A.; HUANG, G.-B. Towards an intelligent framework for multimodal affective data analysis. **Neural Networks**, Elsevier, v. 63, p. 104–116, 2015. Citado nas páginas 36 e 37.
- RAMAKRISHNAN, S.; EMARY, I. M. E. Speech emotion recognition approaches in human computer interaction. **Telecommunication Systems**, Springer, v. 52, n. 3, p. 1467–1478, 2013. Citado nas páginas 35 e 36.
- RIALLE, V.; DUCHENE, F.; NOURY, N.; BAJOLLE, L.; DEMONGEOT, J. Health "Smart"home: information technology for patients at home. **Telemedicine journal and e-health : the official journal of the American Telemedicine Association**, v. 8, n. 4, p. 395–409, 2002. Citado na página 25.
- ROMERO, E.; ARAUJO, A.; MOYA, J. M.; GOYENECHE, J.-M. de; VALLEJO, J. C.; MALLAGON, P.; VILLANUEVA, D.; FRAGA, D. Image processing based services for ambient assistant scenarios. In: **Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living**. [S.l.]: Springer Berlin Heidelberg, 2009, (Lecture Notes in Computer Science, v. 5518). p. 800–807. ISBN 978-3-642-02480-1. Citado nas páginas 25 e 26.
- ROSAS, V. P.; MIHALCEA, R.; MORENCY, L.-P. Multimodal sentiment analysis of spanish online videos. **IEEE Intelligent Systems**, IEEE, v. 28, n. 3, p. 38–45, 2013. Citado na página 36.
- RUSSELL, J. A circumplex model of affect. **Journal of Personality and Social Psychology**, 1980. Citado na página 31.
- SALEM, J. E.; KRING, A. M.; KERR, S. L. More evidence for generalized poor performance in facial emotion perception in schizophrenia. **Journal of abnormal psychology**, American Psychological Association, v. 105, n. 3, p. 480, 1996. Citado na página 26.
- SARKAR, C.; BHATIA, S.; AGARWAL, A.; LI, J. Feature analysis for computational personality recognition using youtube personality data set. In: **ACM. Proceedings of the 2014 ACM multi media on workshop on computational personality recognition**. [S.l.], 2014. p. 11–14. Citado na página 36.
- SAWADA, L. O.; MANO, L. Y.; NETO, J. R. T.; UEYAMA, J. A module-based framework to emotion recognition by speech: a case study in clinical simulation. **Journal of Ambient Intelligence and Humanized Computing**, 2019. ISSN 1868-5145. Disponível em: <<https://doi.org/10.1007/s12652-019-01280-8>>. Citado nas páginas 48, 49, 59 e 63.
- SCHERER, K. R. Appraisal considered as a process of multilevel sequential checking. **Appraisal processes in emotion: Theory, methods, research**, v. 92, p. 120, 2001. Citado nas páginas 31, 33 e 37.
- _____. What are emotions? and how can they be measured? **Social science information**, Sage Publications, v. 44, n. 4, p. 695–729, 2005. Citado nas páginas 16, 31, 32, 33, 34, 37, 41, 51, 67 e 76.

STANKOVIC, J.; CAO, Q.; DOAN, T.; FANG, L.; HE, Z.; KIRAN, R.; LIN, S.; SON, S.; STOLERU, R.; WOOD, A. Wireless sensor networks for in-home healthcare: Potential and challenges. In: **High confidence medical device software and systems (HCMDSS) workshop**. [S.l.: s.n.], 2005. p. 2–3. Citado na página 25.

STEMMLER, G. Methodological considerations in the psychophysiological study of emotion. **Handbook of affective sciences**, p. 225–255, 2003. Citado nas páginas 31 e 33.

SYSTEM, M. A. **QMedic - Medical Alert System - Full Review**. 2015. Disponível em: <<http://medicalalertsystemreviews.net/qmedic-full-review/>>. Citado na página 80.

TAKÁCS, B.; HANÁK, D. A mobile system for assisted living with ambient facial interfaces. **International Journal on Computer Science and Information System**, Citeseer, v. 2, p. 33–50, 2007. Citado na página 30.

TALAL, M.; ZAIDAN, A.; ZAIDAN, B.; ALBAHRI, A.; ALAMOUDI, A.; ALBAHRI, O.; ALSALEM, M.; LIM, C.; TAN, K. L.; SHIR, W. *et al.* Smart home-based iot for real-time and secure remote health monitoring of triage and priority system using body sensors: Multi-driven systematic review. **Journal of medical systems**, Springer, v. 43, n. 3, p. 42, 2019. Citado na página 45.

TAUSCZIK, Y. R.; PENNEBAKER, J. W. The psychological meaning of words: Liwc and computerized text analysis methods. **Journal of language and social psychology**, Sage Publications Sage CA: Los Angeles, CA, v. 29, n. 1, p. 24–54, 2010. Citado na página 63.

TOGEIRO, S. M. G. P.; SMITH, A. K. Diagnostics methods for sleep disorders. **Brazilian Journal of Psychiatry**, SciELO Brasil, v. 27, p. 8–15, 2005. Citado nas páginas 44 e 45.

UEYAMA, J.; VILLAS, L. A.; PINTO, A. R.; GONÇALVES, V. P.; PESSIN, G.; PAZZI, R. W.; BRAUN, T. *et al.* Nodepm: A remote monitoring alert system for energy consumption using probabilistic techniques. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 14, n. 1, p. 848–867, 2014. Citado nas páginas 34 e 35.

VALLABH, P.; MALEKIAN, R. Fall detection monitoring systems: a comprehensive review. **Journal of Ambient Intelligence and Humanized Computing**, Springer, v. 9, n. 6, p. 1809–1833, 2018. Citado na página 42.

WANG, Y.; WANG, S.; TANG, J.; LIU, H.; LI, B. Unsupervised sentiment analysis for social media images. In: AAAI PRESS. **Proceedings of the 24th International Conference on Artificial Intelligence**. [S.l.], 2015. p. 2378–2379. Citado nas páginas 40 e 42.

WÖLLMER, M.; KAISER, M.; EYBEN, F.; SCHULLER, B.; RIGOLL, G. Lstm-modeling of continuous emotions in an audiovisual affect recognition framework. **Image and Vision Computing**, Elsevier, v. 31, n. 2, p. 153–163, 2013. Citado na página 37.

WÖLLMER, M.; WENINGER, F.; KNAUP, T.; SCHULLER, B.; SUN, C.; SAGAE, K.; MORENCY, L.-P. Youtube movie reviews: Sentiment analysis in an audio-visual context. **IEEE Intelligent Systems**, IEEE, v. 28, n. 3, p. 46–53, 2013. Citado na página 36.

XAVIER, R. A. C.; GARCIA, F. E.; NERIS, V. P. de A. Decisões de design de interfaces ruins eo impacto delas na interação: um estudo preliminar considerando o estado emocional de idosos. In: BRAZILIAN COMPUTER SOCIETY. **Proceedings of the 11th Brazilian Symposium on Human Factors in Computing Systems**. [S.l.], 2012. p. 127–136. Citado nas páginas 34 e 37.

YAO, K.-W.; YU, S.; CHENG, S.-P.; CHEN, I.-J. Relationships between personal, depression and social network factors and sleep quality in community-dwelling older adults. **Journal of Nursing Research**, LWW, v. 16, n. 2, p. 131–139, 2008. Citado nas páginas 85 e 88.

YARKONI, T. Personality in 100,000 words: A large-scale analysis of personality and word use among bloggers. **Journal of research in personality**, Elsevier, v. 44, n. 3, p. 363–373, 2010. Citado na página 63.

YUAN, G.; LIM, T. S.; JUAN, W. K.; RINGO, H. M.-H.; LI, Q. A gmm based 2-stage architecture for multi-subject emotion recognition using physiological responses. In: ACM. **Proceedings of the 1st Augmented Human International Conference**. [S.l.], 2010. p. 3. Citado na página 35.

ZHOU, F.; QU, X.; HELANDER, M. G.; JIAO, J. R. Affect prediction from physiological measures via visual stimuli. **International Journal of Human-Computer Studies**, Elsevier, v. 69, n. 12, p. 801–819, 2011. Citado nas páginas 27, 35, 41 e 42.

COMITÊ DE ÉTICA Nº 155/2013

ESCOLA DE ENFERMAGEM DE RIBEIRÃO PRETO DA UNIVERSIDADE DE SÃO PAULO
Centro Colaborador da Organização Mundial da Saúde para o Desenvolvimento da Pesquisa em Enfermagem
Avenida Bandeirantes, 3900 - Ribeirão Preto - São Paulo - Brasil - CEP 14040-902 - Fone: 55 16 3602.3382 - 55 16 3602.3381 - Fax: 55 16 3602.0518
www.eerp.usp.br - eerp@edu.usp.br

COMITÊ DE ÉTICA EM PESQUISA DA EERP/USP

Of.CEP-EERP/USP – 155/2013

Ribeirão Preto, 05 de Junho de 2013.

Prezada Senhora,

Comunicamos que o projeto de pesquisa, abaixo especificado, foi analisado e considerado **APROVADO** pelo Comitê de Ética em Pesquisa da Escola de Enfermagem de Ribeirão Preto da Universidade de São Paulo, em sua 162ª Reunião Ordinária, realizada em 05 de junho de 2013.

Protocolo CAAE: 10551512.1.0000.5393

Projeto: Simulação de alta-fidelidade: validação e aplicação de instrumentos de ensino-aprendizagem para o Brasil.

Pesquisadores: Alessandra Mazzo

Em atendimento à Resolução 196/96, deverá ser encaminhado ao CEP o relatório final da pesquisa e a publicação de seus resultados, para acompanhamento, bem como comunicada qualquer intercorrência ou a sua interrupção.

Atenciosamente,

Prof. Dr. Claudia Benedita dos Santos
Coordenadora do CEP-EERP/USP

Ilma. Sra.
Profa. Dra. Alessandra Mazzo
Departamento de Enfermagem Geral e Especializada
Escola de Enfermagem de Ribeirão Preto - USP