

**INSTITUTO FEDERAL DE MINAS GERAIS
CAMPUS SÃO JOÃO EVANGELISTA**

SÓCRATES ALVES MAGALHÃES; VÂNIA BATISTA DOS SANTOS

**SISTEMA DE APOIO NO PROCESSO DE DIAGNÓSTICO DA DENGUE APLICANDO
APRENDIZADO DE MÁQUINA COM O MÉTODO SUPERVISIONADO**

**SÃO JOÃO EVANGELISTA
2016**

SÓCRATES ALVES MAGALHÃES; VÂNIA BATISTA DOS SANTOS

**SISTEMA DE APOIO NO PROCESSO DE DIAGNÓSTICO DA DENGUE APLICANDO
APRENDIZADO DE MÁQUINA COM O MÉTODO SUPERVISIONADO**

Trabalho de conclusão de curso apresentado ao Instituto Federal de Minas Gerais – *Campus* São João Evangelista como exigência parcial para obtenção do título de Bacharel em Sistemas de Informação.

Orientador: Me. Fábio Rodrigues Martins
Coorientador: Dr. Wesley Gomes de Almeida

SÓCRATES ALVES MAGALHÃES; VÂNIA BATISTA DOS SANTOS

**SISTEMA DE APOIO NO PROCESSO DE DIAGNÓSTICO DA DENGUE APLICANDO
APRENDIZADO DE MÁQUINA COM O MÉTODO SUPERVISIONADO**

Trabalho de conclusão de curso apresentado ao Instituto Federal de Minas Gerais – *Campus* São João Evangelista como exigência parcial para obtenção do título de Bacharel em Sistemas de Informação.

BANCA EXAMINADORA

Orientador: Me. Fábio Rodrigues Martins

Instituição: Instituto Federal de Minas Gerais – *Campus* São João Evangelista

Coorientador: Dr. Wesley Gomes de Almeida

Instituição: Instituto Federal de Minas Gerais – *Campus* São João Evangelista

Convidada: Ma. Márcia Cristina de Paula Cesário

Instituição: Instituto Federal de Minas Gerais – *Campus* São João Evangelista

**SÃO JOÃO EVANGELISTA
2016**

*A Deus que nos capacitou e abençoou na realização deste projeto, rendemos glória e louvores!
Às nossas mães (sempre presentes), fontes de exemplo e inspiração, pois na parede da memória os seus quadros são ilustrados com cores e luzes que nos fortalecem a alma na busca da realização, isto tudo por meio de ensinamentos que de forma perene nos acompanham e iluminam. Essa vitória dedicamos a vocês!*

AGRADECIMENTOS

“Bendirei continuamente ao Senhor, seu louvor não deixará meus lábios” (Sl 33, 2). Graças seriam poucas para ofertarmos a Deus como gratidão pela poderosa ação sobre nossas vidas.

Agradecemos aos nossos familiares e amigos pela presença e apoio constante nessa etapa de nossa trajetória.

Agradecemos a todos os professores e mestres que, não só passaram, mas marcaram-nos por meio de seus talentos com a arte de ensinar.

Em especial, agradecemos ao Professor Me. Fábio Rodrigues Martins que juntamente a nós encarou esse envolvente desafio nos orientando com excelência e dedicação. Ao nosso coorientador Professor Dr. Wesley Gomes de Almeida que de forma corroborativa engrandeceu o trabalho, e a Professora Ma. Marcia Cristina de Paula Cesário que com suas considerações e sugestões proveu suporte para o desenvolvimento do trabalho.

Somos profundamente gratos ao Médico Dr. Marcos Geraldo Godinho Pereira pela atenção e colaboração através de seus conhecimentos, e ainda às Secretarias Municipais de Saúde de Governador Valadares e São João Evangelista pela generosidade que nos atenderam na disponibilização dos dados para a pesquisa; a cada pessoa que de forma particular contribuiu para a concretização deste trabalho, o nosso muito obrigado!

RESUMO

A Inteligência Artificial tem demonstrado substancial importância no desenvolvimento de sistemas para auxiliar os profissionais da área da Medicina e, principalmente, no processo de diagnóstico, no qual a complexidade das variáveis envolvidas dificulta um resultado preciso. O SADDENGUE foi desenvolvido a partir de uma criteriosa análise sobre as manifestações clínicas das pessoas acometidas pela Dengue, uma infecção viral que se não for diagnosticada precocemente pode levar o infectado à morte, e as características em comum que elas apresentam com infectados por outras viroses. O sistema é capaz de apoiar o médico no processo de diagnóstico da Dengue e apresenta uma acurácia de 77,43% na sua classificação. Este trabalho descreve todo o processo de treinamento do sistema, fundamentado pela revisão bibliográfica, a qual objetivamente apresentou as abordagens do Aprendizado de Máquina assim como o contexto atual da Dengue no Brasil. Os dados utilizados no conjunto de treinamento foram provenientes das Secretarias Municipal de Saúde de Governador Valadares e São João Evangelista localizadas em Minas Gerais. Todos os dados advieram de fichas de notificação do agravo da doença e os resultados são comprovados por laboratório. A abordagem de aprendizagem empregada foi o Método Supersessionado e o algoritmo analisado que apresentou a melhor performance foi *Multilayer Perceptron*.

Palavras-chave: Inteligência Artificial. Diagnóstico. Aprendizado de Máquina. Dengue.

ABSTRACT

The Artificial Intelligence has demonstrated substantial importance in the development of systems to support professionals from the Medical field and, mainly, on the diagnostic process, in which the complexity of the variables involved difficult an accurate result. The SADDENGUE was developed from a careful analysis of the clinical manifestations of people affected by dengue, a viral infection that if not diagnosed early can lead the infected to death, and the common characteristics they present with infected by other viruses. The system is able to support the physician in the diagnosis of Dengue process and has an accuracy of 77.43% in its classification. This work describes the entire training process of the system, based on the literature review, which objectively presented Machine Learning approaches as well as the current context of Dengue in Brazil. The data used in the training set were from the Municipal Health Secretaries of Governador Valadares and São João Evangelista located in Minas Gerais. All data were obtained from notification forms of the aggravation of the disease and the results are confirmed by laboratory. The learning approach employed was the Supervised Method and the analyzed algorithm that presented the best performance was Multilayer Perceptron.

Keywords: Artificial Intelligence. Diagnostic. Machine Learning. Dengue.

LISTA DE FIGURAS

Figura 1- As quatro camadas da mineração de dados	19
Figura 2 – Árvore de decisão	20
Figura 3 - Um <i>perceptron</i> simples.....	21
Figura 4 - <i>Perceptron</i> de multicamadas com camadas internas	21
Figura 5 - Os vetores de suporte nas extremidades do hiperplano	22
Figura 6 - Ficha de notificação da Dengue - investigação	31
Figura 7 - Ficha de notificação da Dengue - resultado.....	32
Figura 8 - Distribuição da faixa etária	35
Figura 9 - Distribuição do atributo sexo	35
Figura 10 - Distribuição da classificação	36
Figura 11 - Diagrama de caso de uso	38
Figura 12 - Diagrama de classe	39
Figura 13 - Tela de <i>login</i> inicial	40
Figura 14 - Tela de <i>login</i> com mensagem de erro	40
Figura 15 - Tela de <i>login</i> com direcionamento.....	41
Figura 16 - Tela de entrada de dados.....	41
Figura 17 - Tela de entrada de dados com atributos selecionados	42
Figura 18 - Tela com entrada de dados com <i>feedback</i> sobre o arquivo.....	42
Figura 19 - Tela de retorno do resultado	43
Figura 20 - Tela de informação sobre o sistema.....	43
Figura 21 - Avaliação do algoritmo <i>Multilayer Perceptron</i> aplicando <i>cross-validation</i> - 1	45
Figura 22 - Avaliação do algoritmo <i>Multilayer Perceptron</i> aplicando <i>percentage split</i>	46
Figura 23 - Avaliação do algoritmo J48 aplicando <i>cross-validation</i> - 1	47
Figura 24 - Avaliação do algoritmo SMO aplicando <i>cross-validation</i> - 1	47
Figura 25 - Análise comparativa - primeira etapa	48
Figura 26 - Avaliação do algoritmo J48 aplicando <i>cross-validation</i> - 2	49
Figura 27 - Árvore de decisão gerada pelo algoritmo J48.....	49
Figura 28 - Avaliação do algoritmo SMO aplicando <i>cross-validation</i> - 2	50
Figura 29 - Avaliação do algoritmo <i>Multilayer Perceptron</i> aplicando <i>cross-validation</i> - 2	51
Figura 30 - Avaliação final do <i>Multilayer Perceptron</i>	52
Figura 31 - Análise comparativa final	52

LISTA DE TABELAS

Tabela 1 - Sinais e sintomas	24
Tabela 2 - Sintomas e características.....	33

LISTA DE SIGLAS

CCs – Cardiopatias Congênitas.

CSV – *Coma-separated Values*.

IA – Inteligência Artificial.

IFMG – SJE – Instituto Federal de Minas Gerais – *campus* São João Evangelista.

J48 – *Java 48*.

KDD – *Knowledge Discovery from Data*.

MIT – *Massachusetts Institute of Technology*.

OMS – Organização Mundial de Saúde.

PO – Pesquisa Operacional.

RNA – *Ribonucleic Acid*.

RNA – Redes Neurais Artificiais.

SAD – Sistemas de Apoio à Decisão.

SADC – Sistemas de Apoio à Decisão Clínica.

SADDENGUE – Sistema de Apoio ao Diagnóstico da Dengue.

SE – Sistemas Especialistas.

SMO – *Sequential Minimal Optimization*.

SVM – *Support Vectors Machine*.

UERJ – Universidade Estadual do Rio de Janeiro.

UFPE – Universidade Federal de Pernambuco.

UML – *Unified Modeling Language*.

Weka – *Waikato Environment for Knowledge Analysis*.

SUMÁRIO

1.	INTRODUÇÃO.....	12
1.1	OBJETIVO GERAL.....	14
1.2	OBJETIVOS ESPECÍFICOS	14
2.	FUNDAMENTAÇÃO TEÓRICA.....	15
2.1	SISTEMAS DE APOIO À DECISÃO	15
2.1.1	Sistemas Especialistas	15
2.1.2	Sistemas Baseados no Conhecimento	16
2.2	O APRENDIZADO DE MÁQUINA	16
2.2.1	Mineração de Dados.....	18
2.2.2	Árvores de Decisão	19
2.2.3	Redes Neurais Artificiais	20
2.2.4	Máquina de Vetores de Suporte.....	22
2.2.5	O software Weka	23
2.3	O CONTEXTO DA DENGUE E SUAS CORRELATAS.....	23
2.3.1	Víroses Correlatas.....	25
2.4	TRABALHOS RELACIONADOS	25
3.	METODOLOGIA.....	28
3.1	CLASSIFICAÇÃO DA PESQUISA	28
3.2	DESCRIÇÃO DOS PROCEDIMENTOS ADOTADOS	28
3.3	ESPECIFICAÇÃO DA POPULAÇÃO ENVOLVIDA	29
3.4	TÉCNICAS DE ANÁLISE DOS DADOS	29
3.4.1	Etapas do processo	30
3.5	MODELAGEM DO SOFTWARE.....	37
3.5.1	Diagrama de caso de uso.....	37
3.5.2	Diagrama de classe	38
3.6	O SISTEMA – SADDENGUE.....	39
4.	RESULTADOS E DISCUSSÕES.....	44
3.4	O TREINAMENTO DO SISTEMA	44
3.4.8	Análise dos Algoritmos - Primeira Etapa.....	44
4.1.2	Análise dos Algoritmos - Segunda Etapa	48
5.	CONSIDERAÇÕES FINAIS	53

REFERÊNCIAS	54
ANEXO A – DOCUMENTO DE SOLICITAÇÃO / FRENTE.....	59
ANEXO B – DOCUMENTO DE SOLICITAÇÃO / VERSO	60

1. INTRODUÇÃO

O computador tem sido instrumento de apoio à humanidade desde a sua concepção e o avanço tecnológico, com ênfase nos sistemas computacionais, transformou a rotina das organizações ao passo que agilizou e facilitou seus processos. O homem, com sua inteligência natural, acelerou o processo de desenvolvimento da máquina e criou mecanismos capazes de, artificialmente, reproduzir sua inteligência através dela. Com isso, surgiram os sistemas desenvolvidos através da Inteligência Artificial (IA); conforme Bench-Capon (2015), “Inteligência Artificial é por natureza uma ciência experimental”.

Segundo Bittencourt (2006), a IA é um ramo da ciência da computação com objetivos teórico e prático, no qual o primeiro é referente a criação de modelos cognitivos e o segundo na criação de sistemas computacionais derivados destes modelos. Com isso, ela se assemelha a psicologia que tem como foco a psique humana.

Sistemas especialistas, os quais empregam técnicas da IA, são desenvolvidos e aplicados em diversos setores como o administrativo, educacional, engenharias, médico, militar, ciências aplicadas e muitos outros (CASTILLO, 2012). Eles possuem a capacidade de auxiliar os especialistas humanos nas tomadas de decisões, uma vez que são capazes de armazenar e processar uma quantidade de informações inalcançáveis pela mente humana. Estes sistemas são também denominados sistemas baseados em conhecimento, implementados com informações precisas, ou regras, de uma determinada área específica, transferidas para a máquina através de um especialista (HOFFMAN, 2014).

Diferente dos sistemas especialistas destacam-se os sistemas inteligentes que apresentam um nível mais alto da IA, pois além de serem baseados em conhecimento são capazes de aprender com novas situações e se adaptam a elas aproximando-se da capacidade de aprendizagem humana (ANDRADE, 1999). Conforme Oliveira Junior *et al.* (2007), estes sistemas não invalidam os existentes e nem são a solução para tudo. No entanto, associadas a outras áreas, como a Pesquisa Operacional (PO), são instrumentos poderosos capazes de resolver problemas que desbravam soluções não alcançadas através das técnicas então existentes.

Deste modo, para Wechsler *et al.* (2003), concatenam-se os Sistemas de Apoio à Decisão (SAD), que tem a finalidade geral de auxiliar o ser humano na tomada de decisão. A utilização destes sistemas pode ser observada, principalmente, na área da saúde, a qual está em constante mudança devido às inovações e o surgimento de grandes desafios.

Segundo a Organização Mundial da Saúde, a quantidade de informações na área da saúde dobra a cada três anos e atinge o médico de diversas maneiras, tais como: provendo novos métodos de diagnóstico e terapêutica, através do aparecimento de novos princípios químicos, das inovações da área da biologia molecular e da genética (bioinformática), do surgimento de informações sobre interações de drogas (WECHSLER, *et al.*, 2003).

De acordo com Sabbatini (1993), uma das atividades que exige mais do especialista da área da medicina é o diagnóstico médico, uma vez que este profissional necessita reunir um conjunto complexo de informações sobre o paciente como histórico médico familiar, sintomas recentes e ter conhecimento prévio das possíveis enfermidades associadas aos dados descritos. A não existência de uma doença como uma nova virose, ou mesmo uma derivação de uma doença existente, pode dificultar ou atrasar o processo do diagnóstico. Geralmente em países subdesenvolvidos há uma carência acentuada de profissionais no setor de saúde, em que os médicos precisam, em pouco tempo, atender uma demanda considerável de pacientes. País dentre estes se encontra o Brasil que, de 2002 a 2012, apresentava um déficit de 53 mil médicos, o que levou o governo federal a criar o programa social “Mais Médicos” para amenizar essa defasagem. (MINISTÉRIO DA SAÚDE, 2015).

A Dengue é predominantemente frequente em regiões tropicais e subtropicais e consta como um dos maiores problemas de saúde pública em caráter internacional nas áreas urbanas e semiurbanas. Ela é uma doença de etiologia viral, em que o vírus do complexo Dengue ou DENV é transmitido pelo mosquito do gênero *Aedes*, da família *Flaviviridae*. Não existem, cientificamente, métodos de combate ao vírus, a única forma de erradicação da doença se dá através da exterminação do mosquito (WORLD HEALTH ORGANIZATION, 2011).

Segundo Chaves et al. (2015), o atual cenário brasileiro é extremamente alarmante devido ao grande número de registros epidemiológicos que constam casos do vírus da Dengue que circula por todas as regiões do país. Com novos surtos viróticos que o mesmo transmissor (*Aedes Aegypti*) produz, a preocupação fica redobrada em consequência das duas novas doenças que vêm assolando o território brasileiro, sendo elas: *Chikungunya* e Zika. Os vírus das doenças se distinguem ainda que elas sejam provocadas pelo mesmo transmissor.

De acordo com Teixeira et al. (2009), no século XXI, o Brasil foi considerado o país com maior incidência mundial de Dengue, “isso representou 78% de todos os casos notificados nas Américas e 61% de todos os casos relatados à Organização Mundial de Saúde (OMS).” Esse índice é preocupante para países com mais de um milhão de habitantes como o Brasil, que já ultrapassou os 200 milhões. Os impactos sociais e econômicos são visíveis e a demanda por atendimento médico cada vez está maior.

Para Abe et al. (2012), o diagnóstico precoce da Dengue evitaria agravamentos e casos de óbito, especialmente, relacionados à crianças de 15 anos a menos, que representam 25% dos casos notificados.

A inteligência computacional tem apresentado suma importância ao contribuir para a tomada de decisão no processo de diagnóstico médico. Em consonância com o proposto por Annes e Fortes (2013), o estudo e implementação de um SAD em medicina podem auxiliar os profissionais da saúde e ainda serem usados na área acadêmica como ferramenta educacional.

1.1 OBJETIVO GERAL

Neste trabalho foi realizada uma vasta revisão da literatura com o objetivo geral de aplicar os métodos da IA na análise dos dados de pacientes suspeitos da enfermidade transmitida pelo mosquito *Aedes Aegypti*, a Dengue. A partir disso, desenvolveu-se um SAD para auxiliar no diagnóstico da doença, denominado Sistema de Apoio ao Diagnóstico da Dengue (SADDENGUE). Para teste da ferramenta foram utilizados dados de diagnosticados da doença confirmada do município de Governador Valadares e São João Evangelista, ambos de Minas Gerais. O estudo realizou-se com a orientação e colaboração de um médico clínico-geral da comunidade de São João Evangelista e alcançou uma acurácia, precisão, de aproximadamente 77,5%.

1.2 OBJETIVOS ESPECÍFICOS

Os principais objetivos específicos para a análise e desenvolvimento deste projeto foram os seguintes:

- Analisar as técnicas de aprendizado de máquina;
- Coletar, preparar e analisar os dados para o treinamento do sistema;
- E modelar o problema de diagnóstico da Dengue através da extração do conhecimento dos dados analisados.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 SISTEMAS DE APOIO À DECISÃO

Sol *et al.* (2013) discorrem que sistemas de apoio a decisão, no início dos anos 70, eram conceituados como “sistemas computacionais para dar suporte no processo de tomada de decisão”. O desenvolvimento desses sistemas foi focado no intelecto e partiu do pressuposto de trazer eficácia para área gerencial utilizando-se para isso a tecnologia interativa.

Em meados dos anos 70, o movimento dos SAD reforçou o conceito salientando que estes são, de acordo com Sol *et al.* (2013), “sistemas computacionais interativos, os quais ajudam os tomadores de decisão a utilizarem bases de dados e modelos para resolver problemas mal estruturados. Com isso, o foco ficou definido no suporte e não na tomada de decisão. Contudo, no final dos anos 70 e início dos anos 80, segundo o mesmo autor, esses sistemas passaram a ser descritos como “sistemas que dispõem de tecnologia adequada e acessível para melhorar a eficiência de atividades profissionais e de gerenciamento”. Assim, vários sistemas eram nomeados SAD, e foi nesse período que a psicologia cognitiva e a pesquisa operacional passaram a utilizar essa técnica. Posteriormente, os SAD foram considerados como estações de trabalho inteligentes (SOL *et al.*, 2013).

Dentre a variedade de SAD existentes atualmente, dois exemplos estruturais serão aqui abordados.

2.1.1 Sistemas Especialistas

Os sistemas especialistas (do inglês, *expert systems*), são máquinas que possuem características de um especialista de uma determinada área específica e utilizam um grande volume de dados armazenados do domínio para resolver problemas (MARCUS, 2013).

Estes sistemas recebem todas as informações necessárias para atuarem em uma área particular. A partir do conhecimento, estes devem ser capazes de comunicar com os seres humanos ou com outros sistemas para serem capazes de tomar decisões acertadas em diversas situações, como por exemplo, quando as informações são exatas (precisas) ou incertas (vagas). Sistemas especialistas podem ser entendidos como consultores que auxiliam, ou, em muitos casos, substitui o especialista humano com um moderado grau de confiança. A área da medicina é uma importante aplicação para sistemas especialistas, especialmente no diagnóstico médico, no qual se encontram pacientes com diversas enfermidades que apresentam sintomas variados. Nessa situação, um sistema de apoio ao diagnóstico pode buscar em uma base de dados e extrair

as informações necessárias para auxiliar o médico na tomada de decisão (CASTILLO *et al.*, 2012).

2.1.2 Sistemas Baseados no Conhecimento

Avançando simultaneamente com a tecnologia, os sistemas baseados no conhecimento sucedem os sistemas especialistas. David *et al.* (2012) dissertam que esses sistemas são conhecidos como “segunda geração dos sistemas especialistas”. Embora os autores considerem essa definição, de primeira e segunda geração, um pouco vaga, eles explicam que essa transição se refere à melhores técnicas e à evolução de ideias no desenvolvimento desses sistemas.

*While we cannot give a precise definition of “second generation expert systems”, they have certain common characteristics. First and foremost, is the acknowledgement that knowledge is central in problem solving and that the explicit modeling of knowledge is important for creating understandable and maintainable systems (DAVID *et al.*, 2012).*

Segundo Kruse *et al.* (2012), o objetivo das metodologias baseadas no conhecimento é conseguir uma aproximação entre a codificação, ou representação do conhecimento, através do entendimento dos processos humanos principalmente em cenários de desenvolvimento onde os conceitos, incerteza e imprecisão, são dominantes. Estas técnicas, atualmente, se encontram presentes nas áreas de processamento de imagem, voz, e em robótica, fato que transforma o conceito de IA para além de uma importante ramificação das ciências da computação, sendo então aceita, como um relevante ramo do conhecimento.

2.2 O APRENDIZADO DE MÁQUINA

O processo de aprendizagem envolve uma série de fatores que incluem desde o desenvolvimento de habilidades motoras e cognitivas adquiridas empiricamente, seja por prática ou instrução, à produção do conhecimento advinda de observações e experimentos. Essa capacidade nata de aquisição de conhecimento do ser humano, em suas mais variadas formas, tem despertado nos pesquisadores o desejo de transmiti-la também às máquinas. Alcançar esse objetivo, o aprendizado de máquina (do inglês, “*machine learning*”), e desenvolver modelos computacionais de processos de aprendizagem, tem sido o estudo mais envolvente na área da Inteligência Artificial (MICHALSKI *et al.*, 2013).

Para Bell (2015), aprendizado de máquina é um ramo da IA no qual o computador é usado como ferramenta para treinar sistemas que possam aprender a partir de dados, melhorarem com a experiência e o tempo, e serem capazes de fornecer respostas baseadas na

assimilação de conhecimentos prévios. Essa abordagem divide-se em duas categorias de aprendizagem principais:

- Aprendizado supervisionado; este tipo de aprendizado necessita que o sistema receba instruções, ou seja, ele precisa ser treinado com exemplos. Para cada exemplo os valores devem ser reais; com as entradas X e suas respectivas saídas Y . Com isso, ele deverá ser capaz de classificar novas entradas. Murphy (2012) exemplifica esse tipo de aprendizado com a função:

$$D = \{(x_i, y_i)\}_{i \in N},$$

em que:

- D representa o conjunto de exemplos para treinamento;
- N , o número de exemplos;
- x_i , uma característica ou atributo de uma entidade i ;
- y_i , o resultado ou saída pré-determinada de uma entrada x_i .

Com a utilização desse método, o cuidado com a manipulação dos dados deve ser aplicado, devido a fácil influência que pode ocorrer na seleção destes. E como consequência refletir nos resultados (BELL, 2015).

- Aprendizado não-supervisionado diferentemente do aprendizado supracitado, utiliza um algoritmo que busca por padrões em um conjunto de dados fornecido ao sistema. Não há respostas certas ou erradas, o que se espera como resultado são quais padrões e saídas serão encontrados. Esse tipo de aprendizado costuma ser chamado de “descoberta do conhecimento” e é representado por:

$$D = \{(x_i)\}_{i \in N},$$

em que:

- D representa o conjunto de exemplos para treinamento;
- N , o número de exemplos;
- x_i , uma característica ou atributo de uma entidade i .

Nesse caso, somente são fornecidas as entradas e as saídas são obtidas pelo próprio sistema (MURPHY, 2012).

Aliada às técnicas de aprendizagem então descritas, enquadra-se o aprendizado por reforço. Segundo Lewis e Liu (2013), este modelo de aprendizagem abrange os trabalhos do psicólogo Skinner, teórico influente na abordagem de ensino comportamentalista, o qual desenvolveu estudos que mostraram a mudança de comportamento dos animais em resposta a recompensas e punições.

Esse estilo de aprendizagem apresenta um agente como aprendiz em um ambiente sem que ele receba instruções, as quais o serão passadas através de interações dele com o ambiente.

Sendo assim, ele deverá aprender com os próprios erros e acertos observando as recompensas de suas ações (RUSSEL; NORVING, 2004 *apud* ALVES *et al.*, 2014).

Em sequência serão descritas algumas técnicas empregadas no aprendizado de máquina que foram aplicadas nesse trabalho.

2.2.1 Mineração de Dados

Devido ao crescente volume de dados, estruturados ou não, gerados através da internet e com o uso do computador, tornou-se possível o processo de extração de informações importantes no processo de análise dos dados. O crescimento da mineração de dados nos últimos anos tem sido notório, e essa técnica tornou-se uma disciplina científica no meio acadêmico sendo uma confluência do Aprendizado de Máquina e Estatísticas. Além disso, a mineração de dados tem mostrado sua importância na indústria de computadores sendo, portanto, fundamental na educação em Ciências da Computação (CICHOSZ, 2015).

Larose (2014) aponta para a importância da mineração de dados na significativa concorrência entre as organizações para se manterem no mercado. Aquelas que não utilizam a nova tecnologia acabam perdendo força no mercado e ficam obsoletas. Para ele, a técnica deve ser empregada, pois os analistas humanos não conseguem extrair conhecimento de um vultoso número de dados e, portanto, dentro das organizações a mineração de dados deve ser considerada um processo.

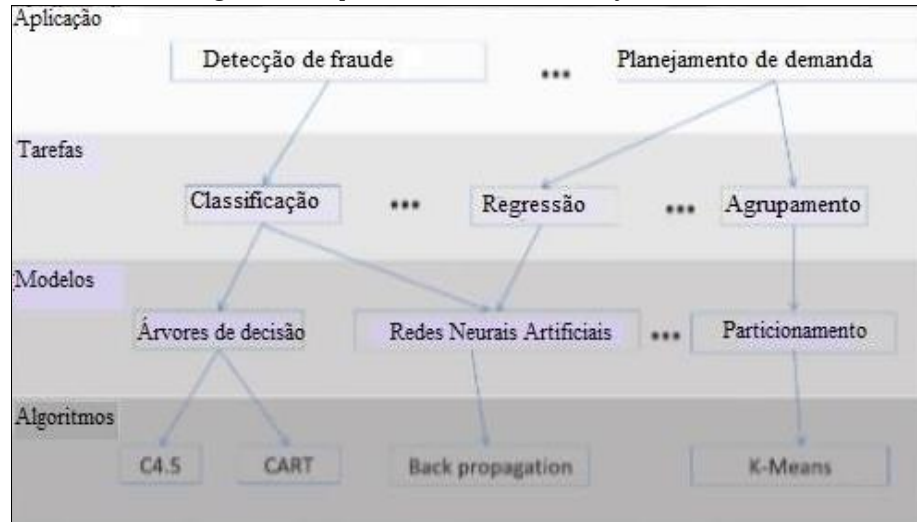
Diante da excessiva quantidade de dados disponíveis em diversas áreas como: científica, médica, financeira, demográfica e comercial, torna-se necessária uma análise automática desses dados, assim como formas de classificá-los e identificar tendências entre eles. Transformar os dados em conhecimento e informação útil são objetivos da mineração de dados que é comumente referida como: descoberta de conhecimento a partir dos dados, do inglês, *Knowledge Discovery from Data* (KDD). Através dela, é possível a extração automática de padrões representando o conhecimento implícito em grandes bases de dados (HAN *et al.*, 2011).

Segundo Han *et al.* (2011), muitos classificam a mineração de dados apenas como parte do processo de descoberta do conhecimento que envolvem sequencialmente, limpeza dos dados, integração, seleção, transformação, a própria mineração, avaliação dos padrões, e apresentação do conhecimento.

Com relação ao processo de mineração, Rokach e Maimon (2014) discorrem sobre a eficiência de se dividir a mineração dos dados em quatro camadas, sendo elas: aplicação, a qual se aplica o problema a ser solucionado; tarefas, que englobam as tarefas de aprendizado de máquina como: classificação, regressão e agrupamento, do inglês *clustering*; a terceira camada

compreende os modelos que podem ser árvores de decisão ou Redes Neurais Artificiais (RNA), no que se refere a classificação, ou particionamento quando se trata de agrupamento; por conseguinte, na quarta e última camada se encontram presentes os algoritmos escolhidos pertencentes a cada modelo. As quatro camadas propostas são representadas através da Figura 1.

Figura 1- As quatro camadas da mineração de dados

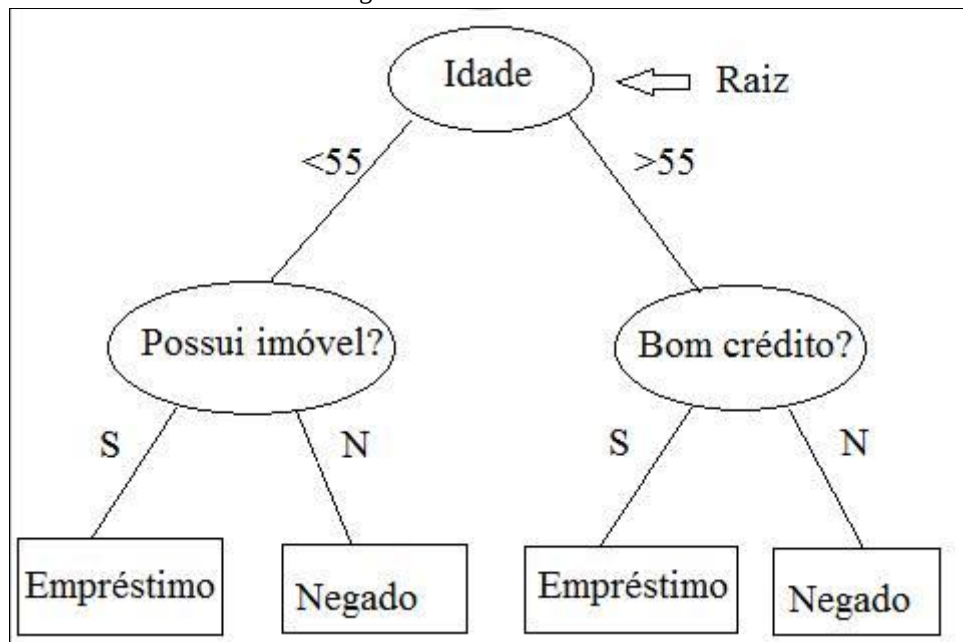


Fonte: Rokach e Maimon (2014). Adaptada pelos autores.

2.2.2 Árvores de Decisão

Segundo Dean (2014), árvores de decisão são técnicas de modelagem de precisão muito conhecidas pela simplicidade e eficácia surpreendente. Embora o método exista desde 1960, a funcionalidade ainda mantém os princípios básicos. Bell (2015) exemplifica o funcionamento de uma árvore de decisão por meio da representação do problema do empréstimo, em que são analisadas condições para chegar a decisão de fornecer, ou não, um empréstimo a um cliente. Para ele, uma árvore é formada por nós, os quais são os valores, ou entradas do problema. Em uma árvore, existem o nó raiz e as folhas, os quais podem ser melhor visualizados na Figura 2. A partir da raiz, indicada pela seta, são colocadas condições e a árvore vai se dividindo até chegar a uma condição de parada, em um dos nós folha, representado pelo retângulo.

Figura 2 – Árvore de decisão



Fonte: Bell (2015). Adaptada pelos autores.

Na Pesquisa Operacional, árvore de decisão é tida como um modelo hierárquico que representa o melhor caminho para a tomada de decisão em um problema específico. Ela pode ser usada para modelos de classificação e regressão, recebendo os referidos nomes quando empregada. É uma potente técnica exploratória, mas não pretende substituir os métodos existentes (ROKACH; MAIMON, 2014).

Quando o objetivo não é apenas criar modelos de classificação eficazes, mas sim fazer uma investigação do modelo, explicar suas previsões e resultados; além de propiciar legibilidade ao modelo, Cichosz (2015) afirma que o método escolhido pela maioria dos mineradores de dados são as árvores de decisão. No entanto, se o objetivo for precisão ao invés de legibilidade, segundo Rokach e Maimon (2014), é aconselhável utilizar as RNA.

2.2.3 Redes Neurais Artificiais

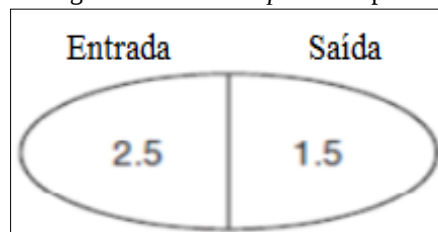
As RNA são modelos computacionais que tem como fundamentação o processo de aquisição de conhecimento que ocorre no cérebro humano, ou seja, o processo que ocorre no sistema nervoso central que envolve os neurônios como principais responsáveis pela recepção, processamento e retransmissão das informações adquiridas através da conexão entre eles (FERREIRA, 2014).

Para Barreto (2002), as RNA apresentam abordagens de soluções para os problemas da IA ao construir sistemas que, semelhantes ao cérebro, possam agir com inteligência. Consequentemente, realizar descobertas e até mesmo aprender coisas novas, sendo, muitas

vezes, imprevisíveis, tal como os animais. As redes neurais são desenvolvidas através de conexões sinápticas entre os neurônios. Assim como acontece no cérebro, existem os neurônios que recebem os sinais externos, como os órgãos do sentido, que nas RNA são as entradas. Os motoneurônios responsáveis por estimular os músculos, que são as saídas, e os neurônios internos ou escondidos. Estes últimos, do inglês *hidden*, são neurônios de suma importância, tanto referindo-se à natureza biológica, quando há a capacidade de desenvolver ações que independem de estímulos externos, quanto à matemática, os quais possibilitam que uma RNA possa resolver problemas mais complexos.

De acordo com Bell (2015), o componente base das RNA é o *perceptron*, o qual, na versão simples, trabalha apenas com números. Ele recebe as entradas, direciona os valores para uma função e retorna o resultado da função (saída). Como exemplificado através da Figura 3.

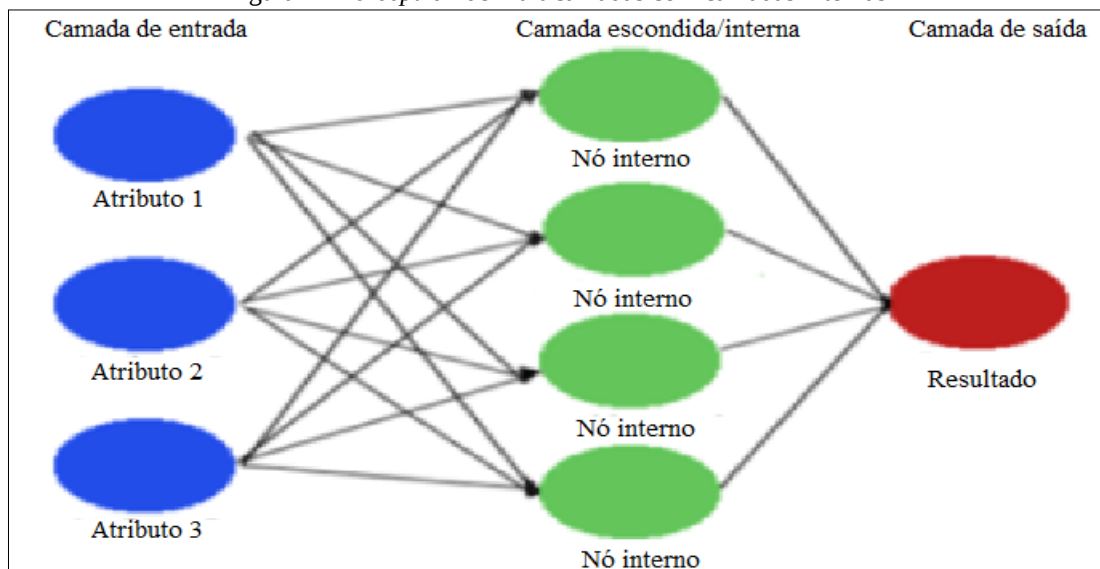
Figura 3 - Um *Perceptron* simples



Fonte: Bell (2015). Adaptada pelos autores.

Os *perceptrons* podem ser de uma ou multicamadas, conforme o que descreve o mesmo autor e ilustrado através da Figura 4, os primeiros são aplicados em problemas linearmente separáveis e não são muito considerados na IA, enquanto os outros, envolvem os neurônios escondidos (*hidden*), citados anteriormente.

Figura 4 - *Perceptron* de multicamadas com camadas internas



Fonte: Ferreira (2014). Adaptada pelos autores.

A camada de entrada contém os atributos provenientes dos dados de treinamento. Em seguida, conectada a ela está a camada oculta ou escondida, a qual é responsável por realizar o processamento das informações de acordo com seus pesos. Por fim, a camada de saída, que retorna o resultado do algoritmo conforme suas entradas (FERREIRA, 2014). O modelo pode ser treinado através do aprendizado supervisionado ou não supervisionado, no entanto, é preferível que se use o primeiro (BELL, 2015).

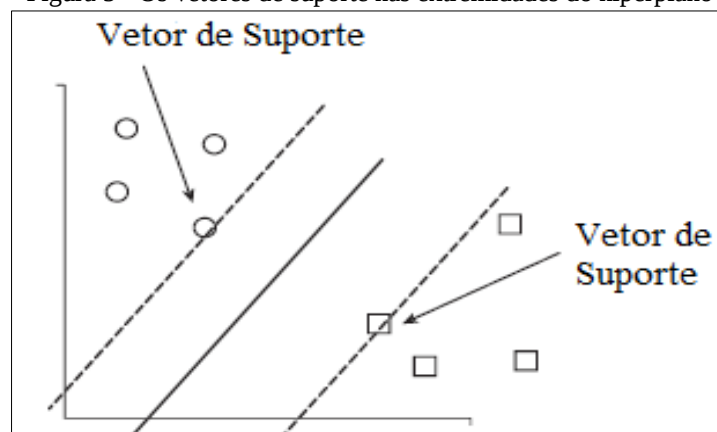
Segundo Dean (2014), em 1989, com o artigo *Multilayer Feedforward Networks are Universal Approximators*, escrito por Kurt Hornik, Maxwell Stinchcombe e Halbert White, ficou provado que através das RNA de três camadas, sendo elas: neurônios de entrada, neurônios internos e neurônios de saída, “seria possível aproximar qualquer função de interesse a qualquer nível de acurácia desejado”. Por isso, as RNA são atualmente os métodos mais utilizados em diversas áreas em que o que se busca é alcançar uma previsão com eficácia.

2.2.4 Máquina de Vetores de Suporte

Do inglês, *Support Vector Machine* (SVM), máquina de vetores de suporte é um algoritmo, dentre os inúmeros existentes para o aprendizado de máquina, que apresenta grande eficiência no modelo de aprendizagem (FERREIRA, 2014). Esse algoritmo objetiva-se criar um hiperplano, linha divisória absoluta, que consiga com eficiência separar dados de entrada em duas classes (se for em uma situação binária), de tal maneira que quando um novo vetor de dados for inserido para teste, seja classificado com acurácia (BELL, 2015).

De acordo com Gonçalves (2015), para que seja encontrado o hiperplano ótimo, é necessário a criação de uma margem máxima entre os exemplos de treinamento, vetores de entrada, e o hiperplano; sendo os vetores que determinam a distância da margem e o hiperplano ótimo, considerados vetores de suporte. Como é representado na Figura 5.

Figura 5 - Os vetores de suporte nas extremidades do hiperplano



Fonte: Bell (2015). Adaptada pelos autores.

Segundo BELL (2015), SVM é uma técnica para classificar objetos e um método supervisionado; ela pode ser comparada com as RNA, pois utilizam alguns métodos em comum, como o conceito do *perceptron*, utilizado, neste caso, para maximizar a margem geométrica entre o hiperplano e os exemplos de treinamento.

2.2.5 O software Weka

O *Weka* (*Waikato Environment for Knowledge Analysis*), nome referente a um pássaro nativo da Nova Zelândia, é um *software* de mineração de dados composto por uma série de algoritmos, os quais são utilizados no processo de aprendizagem de máquina, cuja finalidade consiste em extrair conhecimento a partir de um conjunto de dados. Este *software* foi desenvolvido pelo grupo, que estuda os algoritmos de aprendizagem, na Universidade de *Waikato* em Nova Zelândia. Esta ferramenta possui código aberto, trabalha diretamente com os dados, e por ser desenvolvido em Java, suas bibliotecas podem ser utilizadas em outros projetos desenvolvidos com a mesma linguagem (HALL *et al.*, 2009).

Com *Weka* é possível realizar o pré-processamento dos dados, a seleção das melhores características endereçadas para a busca do conhecimento que se procura descobrir. Ademais, com esse *software* é possível realizar uma classificação dos dados, ao utilizar para isso variados algoritmos de mineração e aprendizado de máquina. Além de ser capaz de avaliar diversos modelos gerados por cada algoritmo (SRIVASTAVA, 2014).

2.3 O CONTEXTO DA DENGUE E SUAS CORRELATAS

A Dengue tornou-se fontes de pesquisas inesgotáveis no Brasil e no mundo, uma vez que os índices de acometimento do vírus cresceram exponencialmente e ainda não possuem tratamentos específicos na contenção desta doença. De acordo com Barreto e Teixeira (2008) existem quatro sorotipos que estão presentes no país: DENV-1, DENV-2, DENV-3, DENV-4.

Classificados assim, de acordo com o tipo de acometimento do complexo Dengue e ao tratar de casos específicos de pessoas infectadas mais de uma vez pelo mosquito, essa natureza dos sorotipos supracitados contribui para a identificação do vírus no organismo, que pode se manifestar tanto a forma clássica quanto na forma mais grave da doença (FIOCRUZ, 2016).

Como apresentado no sétimo parágrafo do capítulo primeiro, a Dengue é uma virose, cujo vírus pertence à família *Flaviviridae*. O Flavivírus apresenta tamanho de 40-50 nm e seu genoma é composto por RNA de fita simples contendo três proteínas estruturais (capsídeo[C], proteína da membrana[M] e glicoproteína do envelope viral[E]), e sete proteínas não estruturais,

as quais são responsáveis pelo processo de replicação do vírus e desencadeamento dos sinais e sintomas no organismo do hospedeiro (SINGHI; KISSOON; BANSAL, 2007).

O ciclo de contaminação da dengue inicia-se quando o mosquito *Aedes Aegypti*, transmissor da doença no Brasil, suga o sangue de uma pessoa infectada. O vírus multiplica-se no intestino médio do contaminado e infecta outros tecidos chegando finalmente às glândulas salivares (FIOCRUZ, 2016).

O Ministério da Saúde considera a condição brasileira em meio aos alarmantes quadros de infecção causados pelo *Aedes Aegypti* como a “tríplice epidemia”, devido aos muitos prováveis casos de Zika, e pelo crescimento exorbitante da Dengue e *Chikungunya* que está alastrando por todo país. Os primeiros ataques da infecção por Zika aconteceram no ano de 2014 em decorrência da presença de estrangeiros no Brasil por ocasião da copa do mundo. Os possíveis casos de Zika comprovados em 2014 comparados aos de 2016, mostram que a situação só piorou em relação aos aproximadamente 91,4 mil prováveis casos de Zika no país, até o mês de fevereiro de 2016 (CACIAN, 2016).

A Dengue é a doença que tem trazido maiores danos ao quadro epidemiológico brasileiro, porém é extremamente importante a defesa contra a *Chikungunya*, devido às debilidades e impotências causadas por ela, bem como a Zika, causadora de deformações neurológicas. Os métodos com maior precisão para interromperem esses históricos são: a exterminação do transmissor (mosquito) e maior investimento em estudos na busca de fórmulas para criação de vacinas que tragam imunização aos vírus relatados (CHAVES *et al.*, 2015).

Abaixo, na Tabela 1 são apresentados os principais sintomas que os vírus produzem no organismo com o grau de sua manifestação representados com o sinal de mais ou com a palavra ausente quando não há manifestação, assim como o que diferencia um do outro:

Tabela 1 - Sinais e sintomas

Sinais e sintomas	Dengue	<i>Chikungunya</i>	Zika
Febre	++++	+++	+++
Mialgia/artralgia	+++	++++	++
Edema das extremidades	Ausente	Ausente	++
Exantema maculopapular	++	++	+++
Dor retroorbital	++	+	++
Hiperemia conjuntival	Ausente	+	+++
Linfadenopatia	++	++	+
Hepatomegalia	Ausente	+++	Ausente
Leucopenia/trombocitopenia	+++	+++	Ausente
Hemorragia	+	Ausente	Ausente

Fonte: Secretaria de Vigilância em Saúde – Ministério da Saúde – volume 46 Nº 26, p.4 – 2015 / Febre pelo o Zika: uma revisão narrativa sobre a doença.

2.3.1 Viroses Correlatas

Na atual conjuntura dos surtos de Dengue espalhados pelo mundo pode-se observar novos registros de virulência que os mosquitos *Aedes Aegypti* e *Aedes Albopictus* da família *Flaviviridae* vêm infectando mais e mais pessoas. Isso pode ser observado com uma abordagem no que diz LUZ *et al.* (2015), quando afirmam que o Zika vírus pode também ser transmitido pelo mosquito, e ainda por relação sexual e transfusão sanguínea:

O vírus Zika é um vírus ARN - ou vírus RNA, ribonucleic acid virus, que tem o ácido ribonucleico como seu material genético. Ele pertence ao gênero *Flavivirus*, família *Flaviviridae*. O genoma consiste em uma molécula de RNA, de cadeia simples e de sentido positivo. Alguns estudos relatam três linhagens principais do ZIKV, uma original da Ásia e duas da África. O ZIKV é transmitido principalmente pelos mosquitos *Aedes Aegypti* e *Aedes Albopictus*. Também existe a possibilidade de transmissão pela via sexual, por transfusão sanguínea e neonatal, embora não se saiba o real protagonismo dessas vias de transmissão na propagação da infecção (LUZ *et al.*, 2015).

O que caracteriza-se como um alerta frente aos registros de casos da Zika são as possíveis causas de Microcefalia encontradas em bebês de mães infectadas pelo ZIKV. Estudos vêm sendo realizados em grandes centros de pesquisas como por exemplo no Departamento de Perinatologia na *Medical University Center*, em Ljubljana, Eslovênia, onde receberam uma gestante oriunda do Brasil, mais precisamente de Natal-RN, região brasileira mais afetada inicialmente. A gestante, possivelmente infectada pelo vírus da Zika, submeteu-se a avaliações médicas e ao final de 32 semanas de gestação comprovaram microcefalia em seu bebê, após acompanhamento irrefutável do caso; com efeito, foi solicitada interrupção da gravidez (MLAKAR *et al.*, 2016).

A *Chikungunya* ou febre *Chikungunya* foi relatada, a priori, na região da Tanzânia no ano de 1950. O vírus é da família *Togaviridae*, do gênero *Alphavirus* e também é transmitido pela fêmea do *Aedes Aegypti* e *Aedes Albopictus*, infectados pelo CHIKV. Da língua africana *makonde*, *Chikungunya* significa: “aquele que é contorcido”, dando principal característica ao que a doença causa no organismo: dores fortíssimas nas articulações devido a artrite. Constatada no Brasil, o primeiro caso da transmissão se deu em meados de setembro de 2014 (CHAVES, *et al.*, 2015).

2.4 TRABALHOS RELACIONADOS

A asserção do presente trabalho é fundamentada em outros que foram realizados com cunho exploratório igual ou semelhante, sistemas especialistas de apoio ao diagnóstico médico. Para enriquecer o trabalho, foram utilizadas as ferramentas necessárias do ramo da inteligência

artificial para a construção de um sistema que auxilia aos serviços médicos, contribuindo para precisão no diagnóstico da Dengue, que vem causando alarde e tendo aumento de casos de forma exorbitante na sociedade brasileira, por isso, o sistema é voltado para ela.

Estudantes da Universidade Nove de Julho de São Paulo desenvolveram um estudo sobre o câncer Colorretal e, frente às dificuldades no diagnóstico dele, atribuíram os recursos da informática na precisão e suporte ao diagnóstico da doença. Desenvolveram um sistema de apoio a decisão, em que através deste incorporaram uma infinidade de informações médicas e uma forma de codificar estratégias de seleção e decisão que retornem respostas que são úteis para profissionais da saúde. Utilizando técnicas de IA, atribuíram as funcionalidades da teoria dos *Rough Sets*, que funciona como um modelo matemático para representação do conhecimento e tratamento de incerteza, muito utilizado em base de dados para aquisição do conhecimento, raciocínio indutivo e descoberta do conhecimento (CONCEIÇÃO; FARIAS; SASSI, 2014).

No artigo “Experiências internacionais de aplicação de sistemas de apoio a decisão clínica em gastroenterologia” elaborado por Hummel *et al.* (2011), são descritas experiências de profissionais da saúde que utilizaram Sistemas de Apoio à Decisão Clínica (SADC) na área de gastroenterologia. Essas experiências são relatadas por meio de documentos publicados no período de 2005 a 2010. Foi dado destaque de 78% dos casos analisados apontando a presença da IA no desenvolvimento desse tipo de sistema, ou seja, o uso de redes Bayesianas, Neurais, Árvores de decisão, e em alguns casos faz referência a melhor precisão do diagnóstico para o auxílio na escolha médica do melhor medicamento a ser indicado ao paciente.

Doutorandos em modelagem computacional pela Universidade Estadual do Rio de Janeiro (UERJ) de Nova Friburgo, desenvolveram uma aplicação para auxílio do profissional médico no diagnóstico de pacientes com suspeita de Alzheimer, a partir das técnicas de mineração de dados. Os dados obtidos para realização da pesquisa foram de pacientes do Centro de Alzheimer e Parkinson no município de Campos dos Goytacazes/RJ, a identidade de cada paciente fora preservada anonimamente. A ferramenta utilizada no processo de mineração de dados foi a Weka, atribuídos seus principais recursos obteve-se informações relevantes, inclusive pela acurácia dos dados predispostos que foi de 73,8%, o que em decorrência à taxa de erros é bem significativo totalizada em 26,2%. Para os pesquisadores algumas informações não contidas nas fichas reunidas dos pacientes acabaram por dificultar na melhora da acurácia dos dados. Relacionadas à etapa de pré-processamento dos dados as técnicas utilizadas dentro da ferramenta supracitada foram: *naïve bayes*, redes bayesianas e árvores de decisão, os pesquisadores, após seus testes, comprovaram que os classificadores de melhor aproveitamento

eram bayesianos, com melhor diagnóstico para doença Alzheimer (MOREIRA, NAMEN, 2016).

Membros da Universidade Federal de Pernambuco (UFPE) descrevem em seu trabalho o desenvolvimento de um aplicativo *mobile* que foi desenvolvido com o objetivo de auxiliar no diagnóstico diferencial de Cardiopatias Congênitas (CCs) nos recém-nascidos. Esse aplicativo foi construído a partir de uma plataforma criada e disponibilizada pelo *Massachusetts Institute of Technology* (MIT). Essa plataforma do MIT é chamada de “*App Inventor*”, por ela é possível a construção de aplicativos em dispositivos móveis utilizando uma plataforma de código aberto para sistemas operacionais Android. Baseando-se nessa plataforma foram desenvolvidos fluxogramas por cardiologistas pediátricos com os principais sinais e sintomas das CCs, isso para que em decorrência das respostas do aplicativo em virtudes dos sinais e sintomas lançados nele, facilitem a compreensão de não especialistas que lidam de forma direta com os neonatos e, com isso, diminui-se as taxas de morbimortalidade neonatal (MOURATO; MOSER; MATTOS, 2013).

3. METODOLOGIA

São descritos no presente capítulo os procedimentos empregados neste trabalho, assim como são apresentadas a caracterização da pesquisa, a população estudada e a definição dos instrumentos e técnicas aplicadas na coleta, preparação e análise dos dados.

3.1 CLASSIFICAÇÃO DA PESQUISA

A gênese do projeto constitui-se com usabilidade de recursos essenciais para efetivação das tarefas; recursos esses dispostos no Instituto Federal de Minas Gerais - *campus* São João Evangelista (IFMG-SJE), na Internet, em livros, artigos e demais trabalhos acadêmicos. Em comum acordo com o exposto, a metodologia desta pesquisa é exploratório-descritiva, com enfoque bibliográfico e documental condizendo com o que Gil (2006) propõe a respeito da revisão na literatura e, utilizando métodos científicos já existentes, afim de explorar a confiabilidade destes para execução do trabalho.

Em razão da categorização desta pesquisa, que é de natureza qualitativa afim de possibilitar a inferência nos dados amalhados e que serão tratados pela subjetividade do sujeito que é o principal fator da pesquisa no espaço natural, em que análises feitas por ele alimentam a sistemática na disponibilização de informações que favorecem aos atributos e variáveis que reunidos deterão as pragmáticas abordagens na realização do projeto e, experimentação louvável das possíveis e necessárias técnicas de estudos (BODGAN e BIKEN, 2003).

3.2 DESCRIÇÃO DOS PROCEDIMENTOS ADOTADOS

O sistema de apoio ao diagnóstico médico foi desenvolvido com o intuito de apoiar o processo de diagnóstico da Dengue. A idealização do projeto concretizou-se com a explícita necessidade de amenizar e auxiliar o trabalho dos profissionais da saúde no momento de realizar o diagnóstico. Para isso, o levantamento dos dados para o estudo procedeu-se através de uma minuciosa pesquisa bibliográfica, a qual, como sugerem Lakatos e Marconi (2010), objetiva-se colocar o pesquisador em contato com o conteúdo existente, presente na literatura, abordado na investigação e permitir que se faça uma inspeção deste. Aspirando assim, a compreensão conceitual dos SAD aliada as melhores técnicas de aprendizado de máquina existentes para o desenvolvimento dos sistemas.

Associadamente à técnica anterior, a obtenção dos dados para a análise experimental foi realizada por intermédio da pesquisa documental, caracterizada por não ter ainda recebido

processo analítico e ser de fonte primária (SÁ-SILVA, ALMEIDA, GUINDANI, 2015). Para aquisição dos dados, elaborou-se um ofício, que encontra-se em anexo, com a solicitação deles. Os dados são advindos de arquivos particulares disponibilizados pelo departamento de Epidemiologia dos municípios de Governador Valadares e São João Evangelista, Minas Gerais, os quais foram tratados diretamente pelos pesquisadores, com a finalidade exclusiva de criar um conjunto de dados sobre pacientes com suspeita de Dengue, diagnosticados ou não com a infecção pelo mosquito *Aedes Aegypti*, mas que apresentaram sintomas relacionados a Dengue. Estes dados, por sua vez, foram usados como exemplos positivos para o treinamento do sistema.

3.3 ESPECIFICAÇÃO DA POPULAÇÃO ENVOLVIDA

Com vistas à criação de um modelo computacional capaz de realizar inferência sobre o diagnóstico da Dengue, a população envolvida é caracterizada pela constatação de sintomas que podem identificar a presença da infecção pelo mosquito *Aedes Aegypti*.

Conforme descrevem Lakatos e Marconi (2010), devido a diversas variáveis que impedem o estudo por completo do universo da pesquisa, este projeto será realizado com uma amostragem representativa da população. Com a finalidade exclusiva de obter dados suficientes para o treinamento do sistema em questão.

O projeto foi realizado com enfoque no município de São João Evangelista com a obtenção de dados de forma preponderante no município de Governador Valadares, o qual dispõe de uma base privilegiada de dados referentes ao estudo, por apresentar alto índice de ocorrências da doença.

3.4 TÉCNICAS DE ANÁLISE DOS DADOS

Em comum acordo com o que discorrem Lakatos e Marconi (2010), a análise dos dados é a parte central de uma pesquisa, pois através dela são gerados os resultados que validam o conhecimento almejado. Esta fase envolve análise e interpretação, sendo a primeira responsável por explicar e a segunda por relacionar os significados encontrados com os objetivos propostos.

Para o tratamento e análise dos dados utilizou-se o *software* de mineração de dados *Weka*, reconhecido por seu poderosíssimo conjunto de algoritmos de treinamento, interface gráfica e ferramentas de visualização (DEAN, 2014).

Essa fase constituiu-se por nove etapas principais que são adotadas no processo de mineração de dados, conforme descrevem Rokach e Maimon (2014), sendo elas:

1. Desenvolvimento da compreensão do domínio da aplicação;

2. Criação de um conjunto de dados;
3. Pré-processamento e a limpeza dos dados;
4. Transformação dos dados;
5. Escolha da tarefa de mineração de dados apropriada;
6. Escolha do algoritmo de mineração;
7. Aplicação do algoritmo escolhido;
8. Avaliação;
9. Emprego do conhecimento descoberto.

3.4.1 Etapas do processo

A compreensão do domínio da aplicação, em relação à primeira etapa, adveio do levantamento bibliográfico, em que foram analisadas as principais técnicas de aprendizado de máquina para a construção do sistema. De forma corroborativa, o auxílio do profissional médico elevou o entendimento sobre a aplicação, por meio de discussões referentes à Dengue e seu diagnóstico.

A segunda etapa, possui caráter determinante na elaboração das principais características da população estudada. Dada a sua relevância, os dados levantados foram provenientes da Gerência de Epidemiologia de Governador Valadares e São João Evangelista. Estes, disponíveis em documentos em formato de notificação dos agravos relacionados à sintomatologia da Dengue. Nestes também, constam os atributos dos pacientes, como: nome (anonimizado pelos fornecedores dos dados), sexo, idade, endereço, manifestações clínicas e outros.

O processo do diagnóstico da Dengue começa com a ficha de notificação e encerra-se com a confirmação laboratorial. Sendo assim, os dados empregados nesta pesquisa foram obtidos por meio de fotografias das notificações, as quais apresentavam todas as informações do processo. Essas fotografias foram feitas pelos profissionais dos setores de epidemiologia supracitados, uma vez que os pesquisadores não poderiam diretamente ter acesso aos documentos, por motivos éticos. Nas Figuras 6 e 7, um modelo das fichas que os pesquisadores dispuseram para a criação do conjunto de dados, no qual é possível observar o tratamento anônimo do diagnosticado.

Figura 6 - Ficha de notificação da Dengue - investigação

Dados clínicos (dengue com complicações, FHD e SCD)

A FHD em geral desenvolve-se entre o 3º e o 5º dia de doença, quando há o recrudescimento da febre. A presença de dor abdominal intensa, hepatomegalia dolorosa, hipotermia com sudorese, letargia/agitação, cianose, urtúrias, hipotensão arterial/postural, vômitos persistentes, manifestações neurológicas são indicadores de que o paciente pode evoluir para FHD ou para um quadro mais grave de dengue.

54 Manifestações Hemorrágicas? 1- Sim 2- Não 3- Ignorado

55 Se sim, quais? 1- Sim 2- Não 3- Ignorado

56 Houve extravasamento plasmático? 1- Sim 2- Não 3- Ignorado

57 Se sim, evidenciado por: 1- Hemococentração 2- Derrames cavitários 3- Hipoproteïnemia

58 Plaquetas (menor) mm³

59 No Caso de FHD/SCD Especificar: 1- Grau I 2- Grau II 3- Grau III 4- Grau IV

60 No Caso de Dengue com complicações, que tipo de complicações? 1- Alterações neurológicas 2- Disfunção cardiorrespiratória 3- Insuficiência hepática 4- Plaquetas < 50.000 mm³ 5- Hemorragia digestiva 6- Derrames cavitários 7- Leucopenia < 1000 8- Não se enquadra nos critérios de FHD

61 Ocorreu Hospitalização? 1- Sim 2- Não 3- Ignorado

62 Data da internação

63 UF

64 Município do Hospital

65 Nome do Hospital

66 (DDD) Telefone

Observações Adicionais

SOLICITADO SOROLOGIA: (X) SIM () NÃO () IGNORADO

Município/Unidade de Saúde: Governador Valadares, E-UF Atalaya

Nome: Edineia R. da Silva Função: medica

Dengue

Sinan NET

Assinatura: [Assinatura]

SVS: 26/09/2006

MANIFESTAÇÕES CLÍNICAS:

☐ Febre ☒ Cefaléia ☒ Dor Retro-orbitária ☒ Artralgia

☐ Dor abdominal ☐ Exantema ☐ Petéquias ☐ Hemorragias diversas

☐ Ext. frias/suores ☐ Prostração ☒ Mialgia ☐ Outros

Fonte: Secretaria de Vigilância Epidemiológica de Governador Valadares.

Figura 7 - Ficha de notificação da Dengue - resultado

SINAN
República Federativa do Brasil
Ministério da Saúde
SISTEMA DE INFORMAÇÃO DE AGRAVOS DE NOTIFICAÇÃO
FICHA DE INVESTIGAÇÃO DENGUE

CASO SUSPEITO: pessoa que viva ou tenha viajado nos últimos 14 dias para área onde esteja ocorrendo transmissão de dengue ou tenha presença de *Ae. aegypti* que apresenta febre, usualmente entre 2 e 7 dias, e apresente duas ou mais das seguintes manifestações: náuseas, vômitos, exantema, mialgias, artralgia, cefaleia, dor retroorbital, petéquias ou prova do laço positiva e leucopenia.

Dados Gerais

1 Tipo de Notificação: 2 - Individual

2 Agravado(a): **DENGUE**

3 Data da Notificação: 17/02/2016

4 UF: 5 Município de Notificação: **Goiás**

6 Unidade de Saúde (ou outra fonte notificador(a)): **GPPI**

7 Data dos Primeiros Sintomas: 10/02/2016

8 Data de Nascimento: 10/01/2006

9 Sexo: 10 Sexo: 11 Sexo: 12 Sexo:

Dados laboratoriais e conclusão

31 Data da Investigação: 17/02/2016

32 Ocupação:

Exame Sorológico (IgM)

33 Data da Coleta: 10/02/2016

34 Resultado: 1 - Reagente 2 - Não Reagente 3 - Inconclusivo 4 - Não Realizado

Exame NS1

35 Data da Coleta:

36 Resultado: 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado

Isolamento Viral

37 Data da coleta:

38 Resultado: 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado

RT-PCR

39 Data da Coleta:

40 Resultado: 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não Realizado

Histopatologia

41 Sorotipo: 1 - DEN 1 2 - DEN 2 3 - DEN 3 4 - DEN 4

42 Resultado: 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado

Imunohistoquímica

43 Resultado: 1 - Positivo 2 - Negativo 3 - Inconclusivo 4 - Não realizado

44 Classificação: 5 - Descartado 10 - Dengue 11 - Dengue com sinais de alarme 12 - Dengue Grave

45 Critério de Confirmação/Descarte: 1 - Laboratório 2 - Clínico-Epidemiológico 3 - Em investigação

Local Provável de Infecção (no período de 15 dias)

46 O caso é autóctone do município de residência? 1 - Sim 2 - Não 3 - indeterminado

47 UF: 48 País:

49 Município: 50 Distrito: 51 Bairro:

52 Doença Relacionada ao Trabalho: 1 - Sim 2 - Não 3 - Ignorado

53 Evolução do Caso: 1 - Cura 2 - Óbito por dengue 3 - Óbito por outras causas 4 - Óbito em investigação 5 - Ignorado

54 Data do Óbito: 55 Data do Encerramento: 27/03/2016

Conclusão

Dengue

Sinan Online: MATERIAL ENVIADO À FUNLO EM

SVS: 11/2/2013

Dados Complementares/ Notificação

SVS: 30/10/2007

Fonte: Secretaria de Vigilância Epidemiológica de Governador Valadares.

Os dados contidos na ficha de notificação da Dengue foram utilizados para criação e treinamento do modelo computacional, incluindo principalmente todos os sintomas que são disponibilizados para que o profissional da saúde notifique-os caso seja um agravamento que o paciente atendido apresente.

Consoante ao Ministério da Saúde (2016), a Dengue manifesta-se primeiramente de forma rápida, com febre entre 39 a 40°C, com duração de dois a sete dias, associados à febre apresentam-se os possíveis sintomas de cefaleia, mialgia, artralgia, prostração, astenia, dor retroorbital, exantema, prurido cutâneo; possibilidade de anorexia, náuseas e vômitos. Como o período febril é dificultoso para associação de doença com diagnóstico preciso, é realizado exame de prova de laço que, com resultado positivo pode indicar Dengue. Leva-se em consideração que o resultado de prova de laço ainda que positivo pode relacionar-se também com outras doenças. Constata-se quadros hemorrágicos moderados e leves como petéquias e sangramento de membrana mucosa; aumento da sensibilidade do fígado decorrente ao período febril acentuado em mais de dois dias.

Os sintomas utilizados na pesquisa após limpeza dos dados são apresentados na Tabela 2, abaixo, com suas principais características.

Tabela 2 - Sintomas e características

Sintomas	Características
Artralgia	Dor nas articulações
Cefaleia	Dor de cabeça intensa
Conjuntivite	Inflamação da membrana transparente que recobre o globo ocular
Dor abdominal	Dor entre a região do tórax e a virilha
Dor retroorbital	Dor ao redor dos olhos
Edema	Acúmulo de líquido entre os tecidos do corpo ou inchaço
Exantema	Erupção na pele ou lesão avermelhada
Extremidades frias e suores	Mãos e pés frios que podem indicar problemas vasculares
Febre	Elevação anormal da temperatura
Hemorragias	Intensa perda de sangue
Mialgia	Dor muscular
Náuseas	Sensação de vômitos
Petéquias	Hemorragias nos vasos sanguíneos apresentadas na pele através de pequenos pontos avermelhados
Prostração	Imobilidade física e psíquica
Prurido cutâneo	Coceira
Vômitos	Expulsão incontrolável de substâncias do estômago pela boca.

Fonte: Própria dos autores.

Com o conjunto de dados criado, possibilitou-se o desenvolvimento da terceira etapa, em que os dados foram pré-processados e retirados da base os melhores atributos. Para isso, contou-se com a ajuda do médico para a seleção e limpeza dos dados, processos esses, que foram automatizados por meio do *software Weka*.

Para a efetivação do pré-processamento foram extraídos os dados que a priori seriam utilizados no treinamento do sistema, tais como: idade, sexo, endereço, gestante ou não, classificação da doença, dados laboratoriais e manifestações clínicas.

Devido a inexistência de grande parte dos dados nas notificações, algumas das informações relacionadas aos atributos pré-selecionados inviabilizou-se o emprego desses na base de dados, o que restringiu a aplicação dos mesmos na aprendizagem do sistema. A seção referente aos dados laboratoriais e conclusão apresenta fragmentos importantes como: exame sorológico e histopatologia. Porém, as notificações não contemplavam essas informações, que são de suma importância para o diagnóstico preciso da doença. O trabalho, então, limitou-se às informações existentes nas notificações.

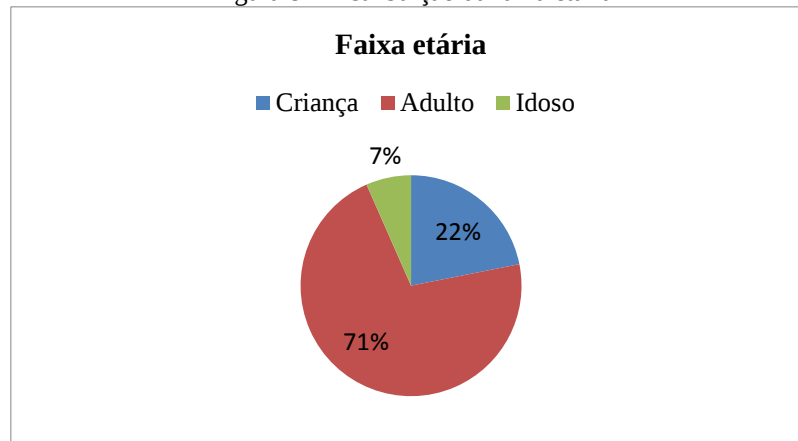
A classificação da doença encontrada em todas as notificações foi referente à Dengue Clássica e casos descartados, ambos laboratorialmente comprovados.

Os dados foram extraídos das fichas e tabulados, manualmente pelos pesquisadores, por meio do software do pacote *office Excel* da Microsoft. O uso da planilha eletrônica contribuiu para a visualização e limpeza dos dados. Alguns desses foram retirados por não colaborarem com a aprendizagem, dentre eles o endereço e instâncias com os mesmos valores atribuídos aos atributos, com a finalidade de obter uma base de dados diversificada composta de bons exemplos.

Para o tratamento no software Weka, é necessário criar um arquivo de texto no formato relação-atributo arff, do inglês *Attribute-Relation File Format*, mas com o arquivo do Excel foi possível convertê-lo em CSV, do inglês *Comma-separated Values*, no qual os valores são separados por vírgula e podem ser usados na ferramenta Weka. De posse do conjunto de dados organizados no arquivo, foi possível acessá-lo na seção de pré-processamento da Weka e visualizar a distribuição de cada atributo e sua relação na base de dados, o que facilitou a escolha dos melhores atributos e aplicação dos filtros necessários.

A transformação dos dados é a parte crucial para a obtenção de bons resultados, por meio dela os dados são organizados de forma que possam ser usados para o treinamento do sistema. A quarta etapa, caracterizou-se pela definição dos atributos e seus valores. Assim sendo, o atributo idade foi transformado para faixa etária para limitar seu valor em três variáveis: criança (de 0-16), adulto (de 17-59) e idoso (maiores que 60). Esse atributo recebeu, respectivamente, valores 1, 2 e 3, os quais eram disponibilizados nas notificações por meio da data de nascimento e para emprego na base de dados, foram transformados pelos pesquisadores. A Figura 8 apresenta como ficou a distribuição da faixa etária, em porcentagem, no conjunto de treinamento final composto por 288 instâncias.

Figura 8 - Distribuição da faixa etária

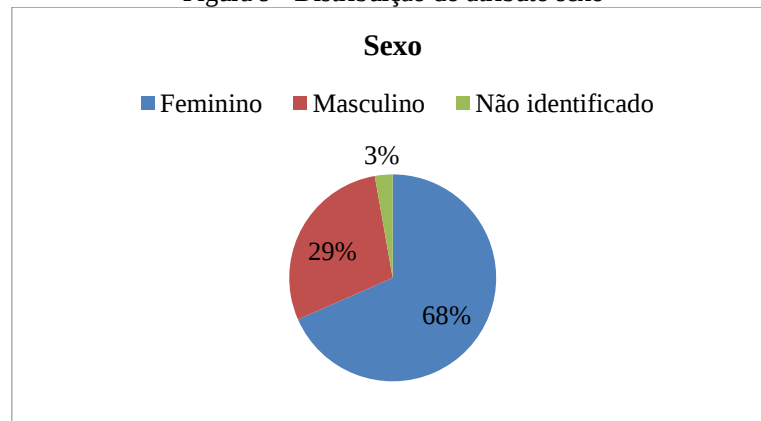


Fonte: Própria dos autores.

Através do atributo faixa etária, buscou-se identificar padrões entre as classes de idades e analisar suas diferenças nas manifestações da doença.

O atributo sexo, utilizado para caracterizar o sujeito, recebeu valores de F para feminino ou M para masculino. O sexo apresentado por cada instância auxiliou na distinção entre as partes e a compreensão de suas características peculiares. Devido à anonimização dos dados pelos fornecedores, em algumas notificações não foi possível identificar o sexo, o qual, na referida situação, recebeu valor em branco, conforme é apresentado na Figura 9.

Figura 9 - Distribuição do atributo sexo



Fonte: Própria dos autores.

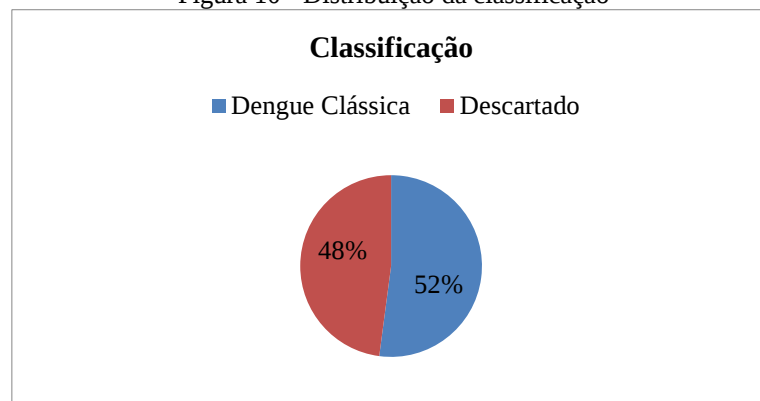
O atributo gestante foi escolhido para entender como a Dengue reage no organismo das grávidas. Esse atributo possui grande relevância por contribuir também para a identificação precoce de possíveis correlatas, como a Zika, que gera preocupação por sua relação com a Microcefalia. Os valores atribuídos a este atributo foram sim, se é gestante, ou não, não é gestante, S ou N respectivamente.

Os outros atributos utilizados são caracterizados pelas manifestações clínicas e informações adicionais apresentadas pelos pacientes, os quais constam na Tabela 2. Todos eles

receberam valores de S para “Sim” (presença dos sintomas) ou N para “Não” (ausência dos sintomas). A escolha desses atributos baseou-se inteiramente nas informações contidas nas notificações.

O atributo classificação determinou a saída, ou seja, a resposta, dadas as entradas. Esse atributo recebeu como valores: “Dengue Clássica” ou “Descartado”, obtidos através da confirmação laboratorial fornecidas nas fichas de notificação. A quantidade de exemplos confirmados e descartados no conjunto de treinamento é representada pelo Figura 10, a qual demonstra como a base de dados alcançada ficou distribuída de forma homogênea.

Figura 10 - Distribuição da classificação



Fonte: Própria dos autores.

Após o conjunto de treinamento do modelo computacional formado, iniciou-se a quinta etapa, na qual foi realizada a escolha da tarefa de mineração de dados. Definida a priori e por se tratar do emprego do método supervisionado, a tarefa foi a classificação, ou seja, diante dos vetores, ou conjunto de valores referentes a cada atributo dos exemplos, apresentados ao sistema, este foi treinado para classificar novos exemplos. Essa fase foi totalmente desenvolvida no *software* Weka, o qual fornece todas as tarefas de mineração com seus respectivos algoritmos de aprendizagem.

A sexta etapa foi expressamente fundamentada pela revisão bibliográfica, a qual abordou os algoritmos de Aprendizado de Máquina que aplicam-se ao uso do método supervisionado. Nessa etapa foram analisados três algoritmos, sendo eles: SMO representando a categoria Máquinas de Vetores de Suporte, J48 com as Árvores de Decisão e *Multilayer Perceptron* aplicando as Redes Neurais Artificiais.

A característica principal da sétima etapa foi a análise constante do algoritmo escolhido. Para isso, as mudanças nos parâmetros aliadas às técnicas de validação disponibilizadas pelo *software* Weka contribuíram para aumentar o desempenho dele.

Na oitava etapa do processo de mineração, os resultados do algoritmo escolhido foram avaliados, interpretados e documentados. Todo o processo de avaliação e interpretação do *Multilayer Perceptron* e os demais algoritmos analisados é descrito no capítulo 4 desse trabalho.

A nona etapa foi a parte final da metodologia adotada, em que o modelo criado foi usado para o desenvolvimento do sistema proposto, o qual é descrito a seguir.

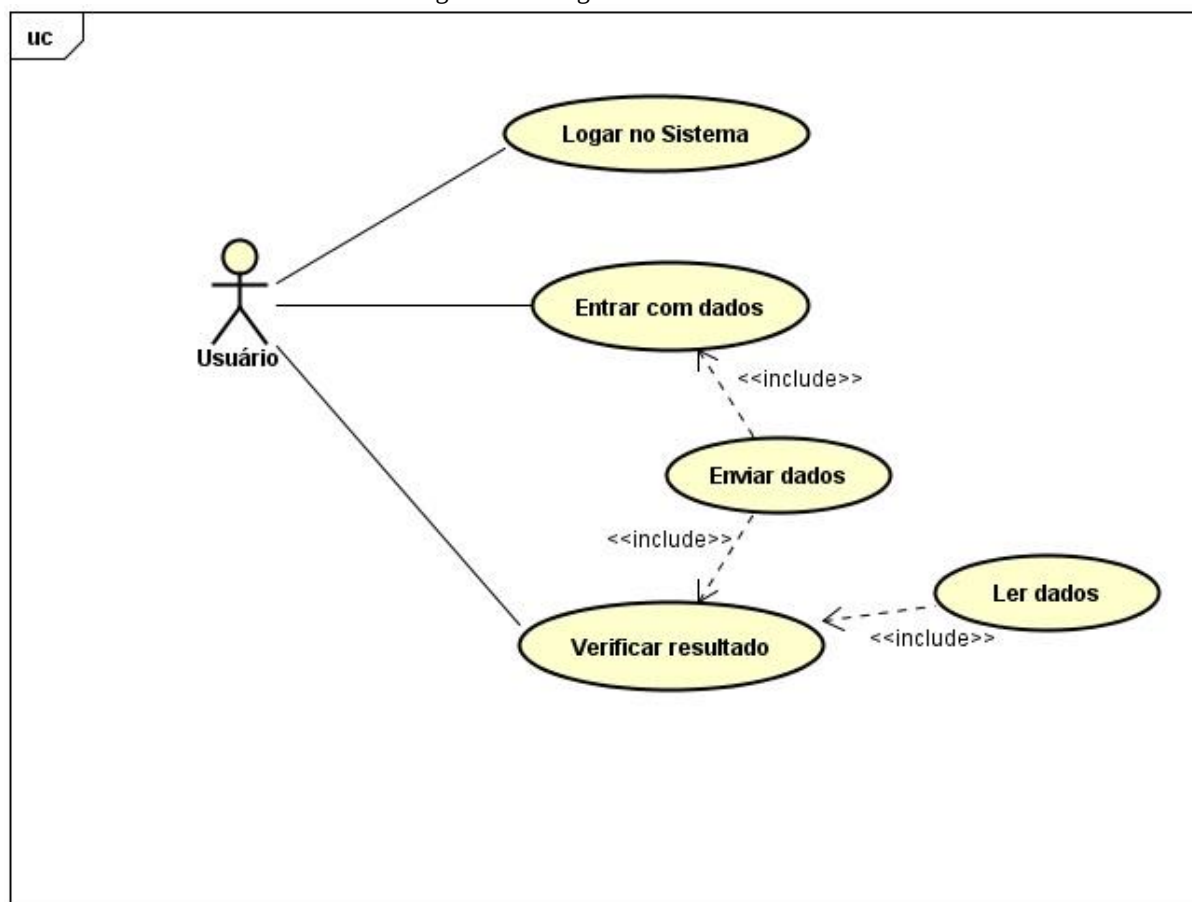
3.5 MODELAGEM DO SOFTWARE

Neste tópico são apresentados os diagramas que foram desenvolvidos para compreensão das funcionalidades e usabilidade do software, fator importante na descrição dos principais componentes abordados nele para demonstração da aplicação da pesquisa realizada para a criação do modelo computacional.

3.5.1 Diagrama de caso de uso

Conforme Sommerville (2008), o diagrama de caso de uso da *Unified Modeling Language* (UML) é um diagrama que é usado para facilitar a comunicação entre os analistas e o cliente. Nele são apresentadas as principais funcionalidades do sistema para o usuário. Abaixo estão listadas as funções do diagrama de caso de uso com a Figura 11.

Figura 11 - Diagrama de caso de uso



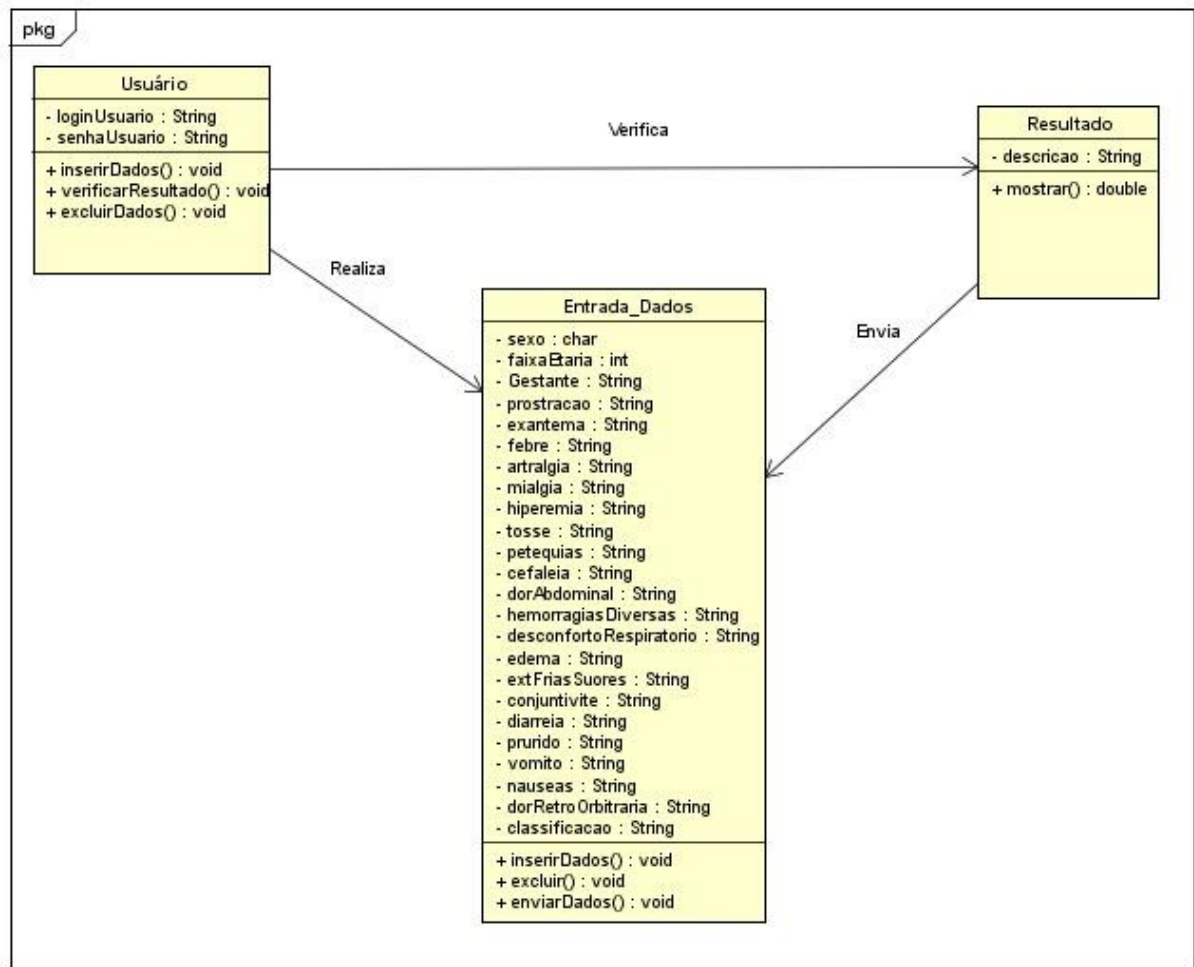
Fonte: Própria dos autores.

Como ilustra a figura superior o usuário poderá logar para efetuar as principais atividades disponibilizadas pelo sistema. Ele poderá realizar as entradas de dados e ainda verificar os resultados a partir do arquivo gerado.

3.5.2 Diagrama de classe

O diagrama de classe é uma representação estática das interações entre as classes do sistema. Ele é utilizado como ferramenta visual para facilitar o trabalho do desenvolvedor. Nele são contidos todos os atributos e métodos pertencentes a cada classe (SOMMERVILLE, 2008). A Figura 12 apresenta as principais classes da modelagem do sistema.

Figura 12 - Diagrama de classe



Fonte: Própria dos autores

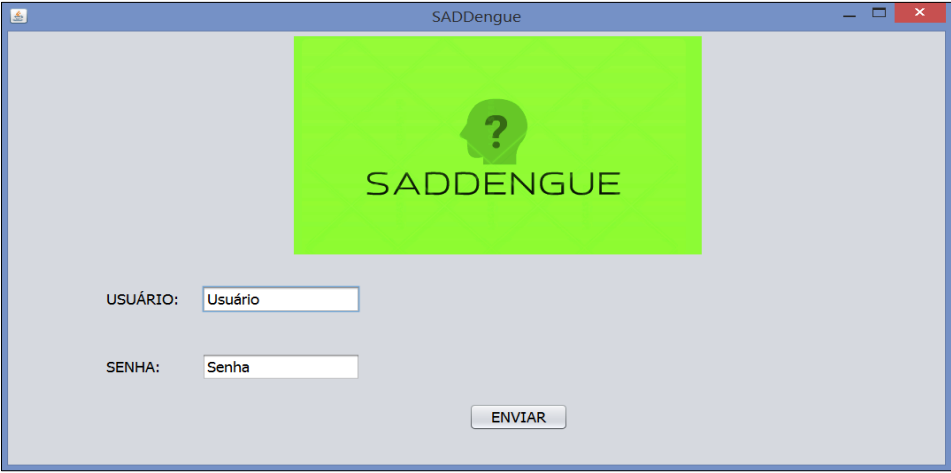
3.6 O SISTEMA – SADDENGUE

Após o modelo computacional criado com o melhor algoritmo escolhido, iniciou-se o desenvolvimento do software. A interface gráfica foi desenvolvida com o intuito de proporcionar ao médico a facilidade de acesso e a simplicidade na interação. Para isso, foram observadas as Heurísticas de Nielsen. Portanto, a tela de entrada dos dados é relativamente parecida com a ficha de notificação da Dengue. O sistema foi desenvolvido na plataforma de desenvolvimento NetBeans e na linguagem de programação Java. A seguir são apresentadas as principais telas que compõem o sistema.

A primeira tela apresentada pela Figura 13, mostra a interface de interação com o sistema e retrata a característica principal dele com a demonstração dos recursos visuais disponibilizados na logomarca, sendo eles: o nome do sistema, SADDENGUE, a imagem que representa a inteligência humana abstraída artificialmente através do símbolo de interrogação e a cor verde que representa a área da Saúde.

A tela de *login* inicial, Figura 13, também contém a opção de que o usuário possa logar no sistema e explorar seus principais recursos. Basta entrar com os dados de usuário e senha e clicar no botão ENVIAR.

Figura 13 - Tela de *login* inicial



Fonte: Própria dos autores.

Após clicar no botão ENVIAR, se o usuário não preencher os campos para realização do *login* corretamente, o sistema retorna uma mensagem informando sobre a obrigatoriedade das informações a serem inseridas de forma correta, conforme apresentado na Figura 14.

Figura 14 - Tela de *login* com mensagem de erro



Fonte: Própria dos autores.

Analisando a Figura 15, com a condição de entrada correta para o acesso ao sistema é devolvida uma mensagem de boas-vindas e clicando no botão OK, este direciona o usuário para a tela principal.

Figura 15 - Tela de login com direcionamento

Fonte: Própria dos autores.

Uma vez que a autenticação aconteceu corretamente, o sistema encaminhará o usuário para a tela principal, apresentada pela Figura 16. Nela são dispostas as entradas de dados relacionadas ao paciente. O médico pode selecionar todas as características e manifestações clínicas apresentadas pelo paciente, conforme na Figura 17.

Figura 16 - Tela de entrada de dados

Fonte: Própria dos autores.

Em seguida ele tem a opção de enviar os dados para avaliação através do botão ENVIAR.

Figura 17 - Tela de entrada de dados com atributos selecionados

The screenshot shows the 'SADDengue' application window with three tabs: 'ENTRADA DE DADOS', 'RESULTADO', and 'O SISTEMA'. The 'ENTRADA DE DADOS' tab is active. It contains three sections: 'DADOS DO PACIENTE', 'MANIFESTAÇÕES CLÍNICAS', and 'DIAGNÓSTICO CLÍNICO'. In the 'DADOS DO PACIENTE' section, 'SEXO' is 'F', 'FAIXA ETÁRIA' is '17-60', and 'GESTANTE' is 'Não'. In the 'MANIFESTAÇÕES CLÍNICAS' section, 'Prostração', 'Exantema', 'Febre', 'Artralgia', 'Cefaléia', 'Desconforto respiratório', 'Prurido', and 'Náuseas' are selected. In the 'DIAGNÓSTICO CLÍNICO' section, 'DENGUE' is selected. An 'ENVIAR' button is at the bottom.

Fonte: Própria dos autores.

Após execução da ação do botão ENVIAR, é gerado um arquivo com todas as informações passadas através da entrada de dados. Este arquivo é gerado em um formato CSV, que pode ser avaliado pelo sistema. Se gerado corretamente o sistema retorna a mensagem de confirmação para o usuário, como apresentada na Figura 18.

Figura 18 - Tela com entrada de dados com feedback sobre o arquivo

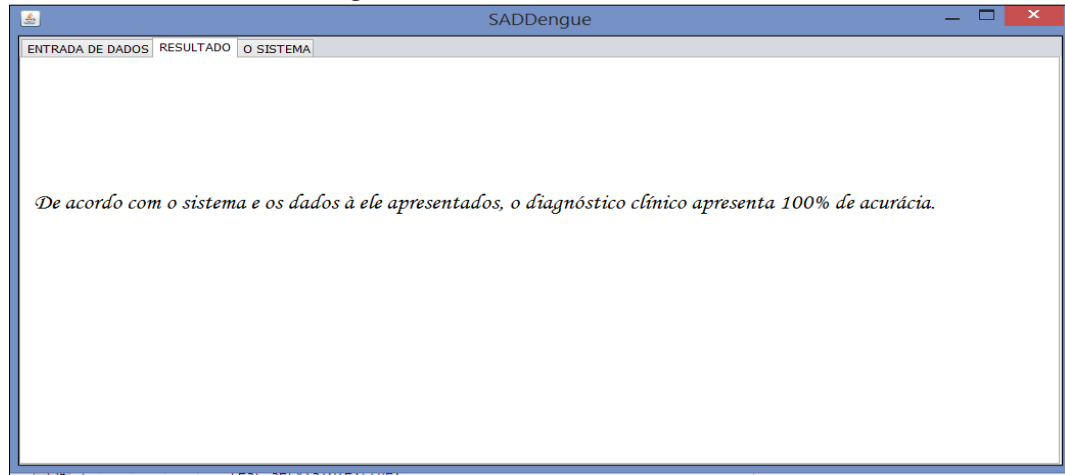
This screenshot is identical to Figure 17, but with a 'Message' dialog box overlaid in the bottom right corner. The message box contains the text 'Arquivo gerado com sucesso!!!' and an 'OK' button. The 'ENVIAR' button in the background form is now disabled.

Fonte: Própria dos autores.

A mensagem retorna que o arquivo foi gerado com sucesso como é retratado na Figura 18. Este arquivo por sua vez, contém todas as seleções aplicadas pelo usuário. Nesse caso, o arquivo apresenta um paciente do sexo feminino, faixa etária entre 17 a 60 anos, adulto, não gestante, encontrando-se prostrada, com febre, exantema, artralgia, cefaleia, prurido, náuseas. O arquivo recebe também a classificação feita pelo diagnóstico clínico, positivo para Dengue.

Após clicar no botão OK, o usuário tem a opção de acessar o resultado da avaliação do sistema, como pode ser visto na Figura 19.

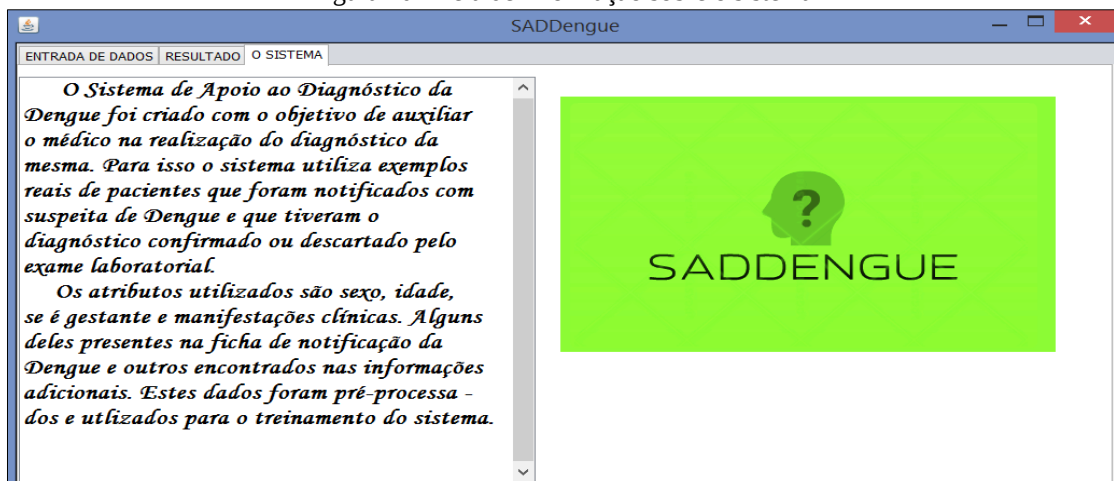
Figura 19 - Tela de retorno do resultado



Fonte: Própria dos autores.

De acordo com as informações contidas no arquivo gerado, o sistema retornou o resultado avaliando a classificação do diagnóstico clínico com 100% de acerto.

Figura 20 - Tela de informação sobre o sistema



Fonte: Própria dos autores.

Ao clicar na aba o sistema, o usuário tem a oportunidade de conhecer um breve relato que traz consigo o objetivo do sistema, como mostra a Figura 20.

4. RESULTADOS E DISCUSSÕES

Com o desenvolvimento do Sistema de Apoio ao Diagnóstico da Dengue percebeu-se a importância de integrar os sistemas computacionais com a área da saúde. Por meio da análise dos dados, foi constatado como o processo de diagnóstico desta virose torna-se um procedimento difícil devido às diversas similaridades entre os sintomas apresentados, os quais também são diretamente relacionados às outras viroses. A priori o objetivo do projeto era criar um sistema que auxiliasse no diagnóstico da Dengue, Zika e Chikungunya. Contudo, pelo fato do surgimento das duas últimas ainda estar recente no Brasil, a disponibilidade de dados, na região, comprovados para a análise impossibilitou a investigação.

A pesquisa iniciou-se com uma busca sistêmica de informações sobre a Dengue, suas causas, a maneira como o vírus infecta e age no organismo humano, assim como as consequências que podem ocorrer não havendo um diagnóstico preciso. A colaboração de um médico clínico-geral de São João Evangelista foi fundamental para o esclarecimento de incógnitas que surgiram no decorrer do trabalho.

Foram analisadas técnicas da IA que melhor pudessem atender o objetivo desse estudo e a fundamentação teórica possibilitou encontrar as melhores, as quais foram empregadas neste trabalho. Diante disso, após a análise dos algoritmos, o *Multilayer Perceptron* foi o escolhido para a criação do modelo computacional.

3.4 O TREINAMENTO DO SISTEMA

O levantamento dos dados para o estudo foi realizado em duas etapas: na primeira foram disponibilizadas 200 instâncias e, a partir delas, foi selecionada, criteriosamente, uma amostra de 127 instâncias; a segunda etapa contou com o adicional de 200 instâncias, analisadas, tratadas e após o tratamento resultou em um total de 288 instâncias na base final do conjunto de treinamento.

3.4.8 Análise dos Algoritmos - Primeira Etapa

Devido ao número de exemplos na primeira etapa aliado à sua complexidade, os resultados obtidos foram insatisfatórios, uma vez que a aprendizagem almejada não foi alcançada com êxito, embora tenha resultado em uma classificação acima de 52% de desempenho em todos os algoritmos testados, em seus melhores resultados. As figuras a seguir apresentam os resultados gerados através do software Weka, da primeira avaliação dos

algoritmos: *Multilayer Perceptron*, representando as Redes Neurais Artificiais; Java 48 (J48), Árvore de Decisão e *Sequential Minimal Optimization* (SMO), Máquina de Vetores de Suporte.

Figura 21 - Avaliação do algoritmo Multilayer Perceptron aplicando cross-validation - 1

```
Time taken to build model: 0.52 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

Correctly Classified Instances	79	62.2047 %
Incorrectly Classified Instances	48	37.7953 %
Kappa statistic	0.2429	
Mean absolute error	0.4785	
Root mean squared error	0.4877	
Relative absolute error	95.6511 %	
Root relative squared error	97.492 %	
Total Number of Instances	127	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.524	0.281	0.647	0.524	0.579	0.247	0.624	0.653	DENGUE CLÁSSICA
	0.719	0.476	0.605	0.719	0.657	0.247	0.624	0.573	DESCARTADO
Weighted Avg.	0.622	0.379	0.626	0.622	0.618	0.247	0.624	0.613	

```
=== Confusion Matrix ===
```

```

a  b  <-- classified as
33 30 | a = DENGUE CLÁSSICA
18 46 | b = DESCARTADO
```

Fonte: Software Weka.

A Figura 21 apresenta a base de dados formada na primeira etapa com um total de 127 exemplos, o algoritmo *Multilayer Perceptron* classificou corretamente 79 exemplos e incorretamente os restantes 48. No processo de treinamento, foi aplicada a técnica chamada de *cross-validation* com 10 *folds*. Por meio dessa abordagem é possível separar o conjunto de treinamento em partes menores, *folds*, e realizar o treinamento ao mesmo tempo que o testa, uma validação cruzada. Essa técnica tem uma característica importante, pois não permite que haja manipulação das amostras para o teste. O sistema se encarrega de, aleatoriamente, escolher os exemplos que servirão de teste, e a função do pesquisador é analisar qual o número de *folds*, desempenhará um melhor resultado.

A avaliação do *Multilayer Perceptron* apresentou, neste primeiro caso, com o uso dos parâmetros padronizados, uma acurácia de 62.20% e uma taxa de erro de 37.79%. A matrix de erro ou *confusion matrix* indica a classificação mostrando o erro em números, o que visualmente facilita o entendimento da performance do algoritmo, como é mostrado na Figura 21.

Na sua melhor performance da primeira etapa, o *Multilayer Perceptron* apresentou 65.11% de acurácia. Para a obtenção desse resultado utilizou-se a técnica *Percentage Split*, por meio da qual o conjunto de treinamento é dividido de acordo com a porcentagem fornecida ao sistema. Por padrão utilizou-se dividir o conjunto em 66%, o que implica que a base se dividiu em 84 exemplos para o treinamento do modelo e 43 para o teste, conforme mostra a Figura 22.

Figura 22 - Avaliação do algoritmo *Multilayer Perceptron* aplicando *percentage split*

```

Time taken to build model: 0.56 seconds

=== Evaluation on test split ===

Time taken to test model on training split: 0 seconds

=== Summary ===

Correctly Classified Instances      28          65.1163 %
Incorrectly Classified Instances    15          34.8837 %
Kappa statistic                    0.2997
Mean absolute error                 0.4876
Root mean squared error             0.4908
Relative absolute error             97.5142 %
Root relative squared error         98.1635 %
Total Number of Instances          43

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.571   0.273   0.667     0.571   0.615     0.303   0.641    0.626   DENGUE CLÁSSICA
                0.727   0.429   0.640     0.727   0.681     0.303   0.641    0.666   DESCARTADO
Weighted Avg.   0.651   0.352   0.653     0.651   0.649     0.303   0.641    0.647

=== Confusion Matrix ===

  a  b  <-- classified as
12  9  | a = DENGUE CLÁSSICA
 6 16 | b = DESCARTADO

```

Fonte: *Software Weka*.

O segundo algoritmo analisado, a árvore de decisão J48, apresentou uma performance bem inferior ao *Multilayer Perceptron* com acurácia de 52.75%, e foram utilizadas as mesmas técnicas para o processo de aprendizagem do sistema. A Figura 23 retrata o resultado da melhor avaliação do algoritmo com a primeira base de dados.

Figura 23 - Avaliação do algoritmo J48 aplicando *cross-validation* - 1

```

Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      67           52.7559 %
Incorrectly Classified Instances    60           47.2441 %
Kappa statistic                    0.0534
Mean absolute error                 0.4633
Root mean squared error             0.5684
Relative absolute error             92.6175 %
Root relative squared error         113.6297 %
Total Number of Instances          127

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.413    0.359    0.531      0.413    0.464      0.055    0.574    0.542     DENGUE CLÁSSICA
                0.641    0.587    0.526      0.641    0.577      0.055    0.574    0.572     DESCARTADO
Weighted Avg.   0.528    0.474    0.528      0.528    0.521      0.055    0.574    0.557

=== Confusion Matrix ===
  a  b  <-- classified as
26 37 | a = DENGUE CLÁSSICA
23 41 | b = DESCARTADO

```

Fonte: *Software Weka*.

A Figura 24 representa a avaliação do terceiro algoritmo analisado, o SMO, que apresentou uma performance melhor que J48, mas não teve um desempenho maior que o *Multilayer Perceptron*. A maior acurácia do algoritmo foi de 59.05%.

Figura 24 - Avaliação do algoritmo SMO aplicando *cross-validation* - 1

```

Time taken to build model: 0.06 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      75           59.0551 %
Incorrectly Classified Instances    52           40.9449 %
Kappa statistic                    0.1796
Mean absolute error                 0.4094
Root mean squared error             0.6399
Relative absolute error             81.8564 %
Root relative squared error         127.9171 %
Total Number of Instances          127

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.476    0.297    0.612      0.476    0.536      0.184    0.590    0.551     DENGUE CLÁSSICA
                0.703    0.524    0.577      0.703    0.634      0.184    0.590    0.555     DESCARTADO
Weighted Avg.   0.591    0.411    0.594      0.591    0.585      0.184    0.590    0.553

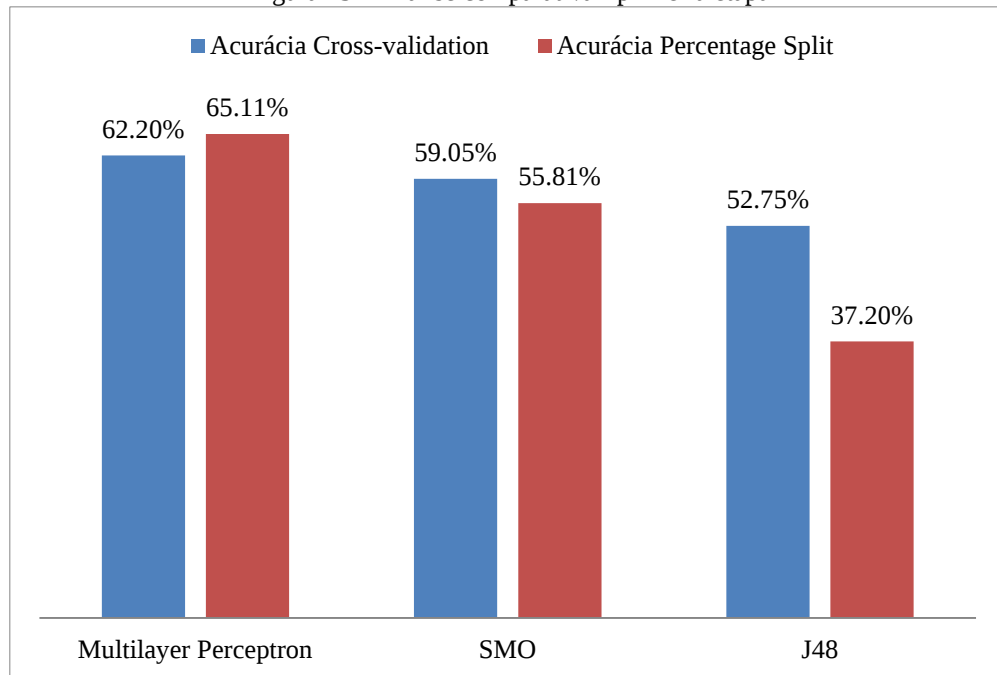
=== Confusion Matrix ===
  a  b  <-- classified as
30 33 | a = DENGUE CLÁSSICA
19 45 | b = DESCARTADO

```

Fonte: *Software Weka*.

A Figura 25 representa a análise dos três algoritmos empregados na classificação e seu desempenho de acordo com a técnica de divisão da base de treinamento aplicada para teste.

Figura 25 - Análise comparativa - primeira etapa



Fonte: Software Weka.

4.1.2 Análise dos Algoritmos - Segunda Etapa

Após o emprego de diversas abordagens no processo de aprendizagem como as técnicas descritas anteriormente, para aprimorar o desempenho do modelo computacional, optou-se por adicionar exemplos à base de treinamento. A segunda etapa do treinamento introduziu-se com um total de 288 instâncias, o conjunto de dados anterior somado ao novo.

A criação do novo conjunto de treinamento viabilizou uma amostragem mais representativa da população estudada, pois permitiu uma diversificação dos exemplos apresentados aos algoritmos. Desta forma, o aprendizado evoluiu e os algoritmos apontaram melhores resultados.

O algoritmo J48 e o SMO apresentaram, em seus melhores resultados, a mesma acurácia de 77.08%, e o *Multilayer Perceptron* alcançou a melhor performance e apresentou 77.43%. Diante da comparação analítica entre os algoritmos estudados, o algoritmo escolhido para a implementação da ferramenta foi o *Multilayer Perceptron*. A avaliação completa dos algoritmos da segunda etapa do treinamento é retratada conforme as figuras 26-31.

Figura 26 - Avaliação do algoritmo J48 aplicando *cross-validation* - 2

```
Time taken to build model: 0.02 seconds
```

```
=== Stratified cross-validation ===
```

```
=== Summary ===
```

Correctly Classified Instances	222	77.0833 %
Incorrectly Classified Instances	66	22.9167 %
Kappa statistic	0.5379	
Mean absolute error	0.2817	
Root mean squared error	0.4193	
Relative absolute error	56.4346 %	
Root relative squared error	83.9279 %	
Total Number of Instances	288	

```
=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.853	0.319	0.744	0.853	0.795	0.544	0.811	0.761	DENGUE CLÁSSICA
	0.681	0.147	0.810	0.681	0.740	0.544	0.811	0.785	DESCARTADO
Weighted Avg.	0.771	0.236	0.776	0.771	0.769	0.544	0.811	0.772	

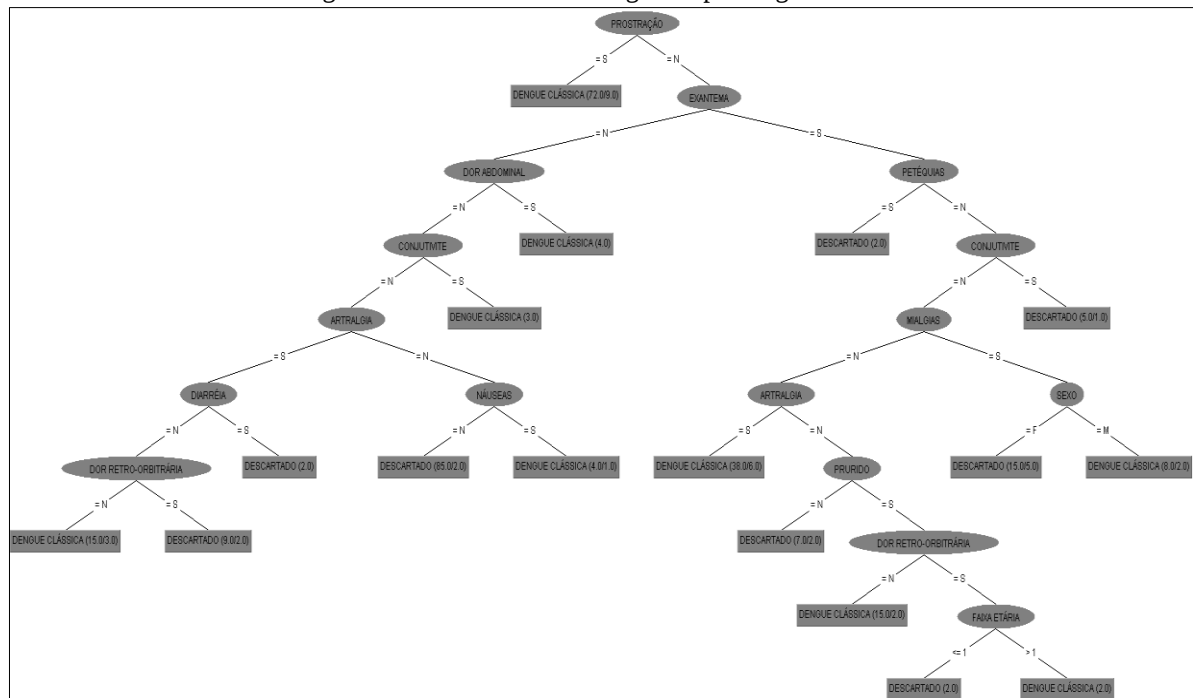
```
=== Confusion Matrix ===
```

a	b	<-- classified as
128	22	a = DENGUE CLÁSSICA
44	94	b = DESCARTADO

Fonte: Software Weka.

A árvore de decisão J48 melhorou seu desempenho de 52.75% para 77.08%, com a inclusão de novos exemplos na base, como é demonstrado por meio da Figura 26, e a representação gráfica da árvore pode ser visualizada na Figura 27, onde é explicitado o atributo mais relevante para a decisão na raiz da árvore.

Figura 27 - Árvore de decisão gerada pelo algoritmo J48



Fonte: Software Weka.

A prostração foi o atributo mais relevante para o aprendizado do sistema, e suas características, fraqueza e moleza corporal, são elas que definem o nome da Dengue, em que a pessoa infectada fica debilitada fisicamente.

O algoritmo SMO também obteve um aumento significativo no seu desempenho com o acréscimo de instâncias ao conjunto de treinamento. Ele avançou de 59.05% para 77.08%, o mesmo desempenho do algoritmo J48, conforme representado a seguir.

Figura 28 - Avaliação do algoritmo SMO aplicando cross-validation - 2

```
Time taken to build model: 0.1 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      222           77.0833 %
Incorrectly Classified Instances    66           22.9167 %
Kappa statistic                    0.5382
Mean absolute error                 0.2292
Root mean squared error             0.4787
Relative absolute error             45.9114 %
Root relative squared error         95.8236 %
Total Number of Instances          288

=== Detailed Accuracy By Class ===
```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.847	0.312	0.747	0.847	0.794	0.544	0.768	0.712	DENGUE CLÁSSICA
	0.688	0.153	0.805	0.688	0.742	0.544	0.768	0.704	DESCARTADO
Weighted Avg.	0.771	0.236	0.775	0.771	0.769	0.544	0.768	0.708	

```

=== Confusion Matrix ===
  a  b  <-- classified as
127 23 |  a = DENGUE CLÁSSICA
 43 95 |  b = DESCARTADO

```

Fonte: Software Weka.

A melhor performance foi apresentada pelo *Multilayer Perceptron* com 77,43% de acurácia. A taxa de erro, a qual representa o número de instâncias do conjunto de treinamento que foram classificadas incorretamente, resultou em 22,56%. Diante dos resultados o algoritmo foi o escolhido para modelar o problema.

Figura 29 - Avaliação do algoritmo Multilayer Perceptron aplicando cross-validation - 2

```

Time taken to build model: 1.67 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      223           77.4306 %
Incorrectly Classified Instances    65           22.5694 %
Kappa statistic                    0.5456
Mean absolute error                 0.2984
Root mean squared error             0.3994
Relative absolute error             59.7793 %
Root relative squared error         79.9559 %
Total Number of Instances          288

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0.840    0.297    0.754      0.840    0.795      0.549    0.843    0.816    DENGUE CLÁSSICA
                0.703    0.160    0.802      0.703    0.749      0.549    0.843    0.836    DESCARTADO
Weighted Avg.   0.774    0.231    0.777      0.774    0.773      0.549    0.843    0.826

=== Confusion Matrix ===

  a  b  <-- classified as
126 24 |  a = DENGUE CLÁSSICA
 41 97 |  b = DESCARTADO

```

Fonte: Software Weka.

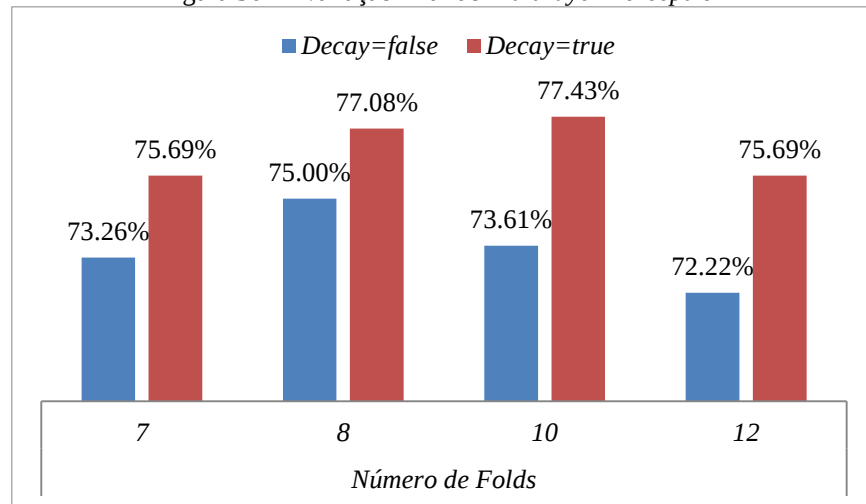
A Figura 29 representa a avaliação final do *Multilayer Perceptron* em sua melhor performance. A técnica aplicada foi a validação cruzada com 10 *folds* e o resultado obtido adveio das mudanças realizadas nos parâmetros padrão do algoritmo.

O processo de treinamento do modelo aplicando o *Multilayer Perceptron*, com seus parâmetros padronizados e a técnica da validação cruzada com o número de *folds* variando de 7 à 12, é demonstrada na Figura 30.

A mudança realizada no parâmetro *decay*, que controla o decaimento da taxa de aprendizagem no momento do treinamento, que por padrão recebe *false*, foi fundamental para alcançar melhores resultados. Esse parâmetro permite que a taxa de aprendizagem decaia e define qual deve ser a taxa atual de aprendizagem no momento do treinamento. Quando mudado para *true*, ou verdadeiro, melhora a performance do algoritmo.

Foram feitas mudanças nos outros parâmetros do algoritmo, mas estas não foram mantidas por não surtirem efeito na avaliação final.

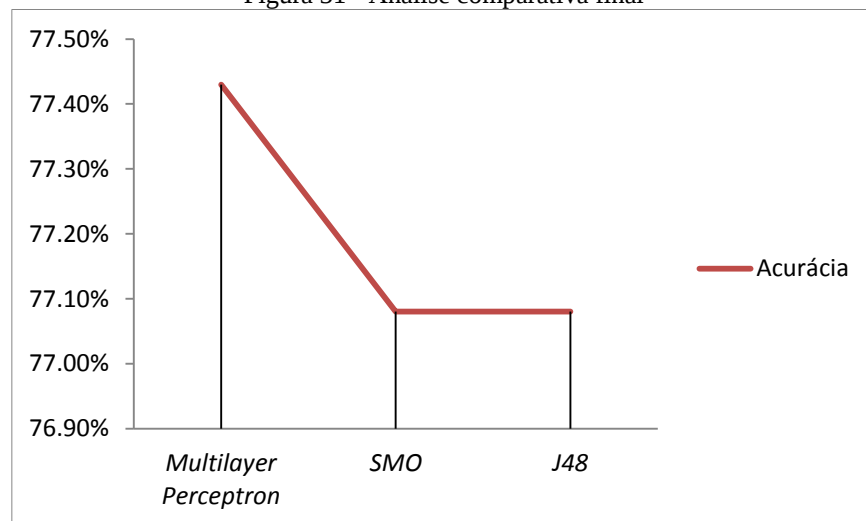
Os resultados da avaliação do *Multilayer Perceptron* após a mudança do parâmetro *decay* podem ser comparados a seguir.

Figura 30 - Avaliação final do *Multilayer Perceptron*

Fonte: Software Weka.

A Figura 31 apresenta a análise comparativa dos três algoritmos assim como representa a segunda etapa do processo de aprendizagem com o conjunto de treinamento final.

Figura 31 - Análise comparativa final



Fonte: Software Weka.

Em relação à pesquisa desenvolvida, o modelo computacional atende de forma satisfatória à aplicação. O sistema criado alcançou o objetivo almejado nesse trabalho, o qual através de seus componentes explicitou de forma simples, clara e concisa o auxílio ao médico no processo de diagnóstico da Dengue.

O SADDENGUE é um modelo simples do que se espera de um SAD; contudo, sua implementação é uma base para o desenvolvimento de um sistema robusto na área de diagnóstico com os irrefutáveis recursos da Inteligência Artificial.

5. CONSIDERAÇÕES FINAIS

A Ciência da Computação, especialmente a área da IA, têm a contribuir com as organizações por meio do apoio de seus sistemas ao desenvolvimento estratégico delas. Com a realização desse trabalho ficou demonstrada a relevância da automatização no processo de análise dos dados para a tomada de decisão no momento de realizar o diagnóstico.

As ferramentas disponíveis para o tratamento dos dados são possuidoras de grandiosa potência e a massante produção de dados auxilia na formação de ricas bases de dados para estudo. Principalmente na área da saúde, os dados produzidos detêm a resposta para muitos problemas da atualidade. A fundamentação teórica foi crucial para a investigação das melhores técnicas de aprendizado de máquina e as ferramentas utilizadas contribuíram para a realização de todo o processo, desde a compreensão do domínio da aplicação, tabulamento e pré-processamento dos dados até a fase de testes na ferramenta desenvolvida.

A Dengue, suas correlatas e as outras viroses existentes, compartilham características semelhantes, o que torna o seu processo de diagnóstico complexo. O profissional da saúde necessita fazer uma cautelosa investigação e as vezes solicita uma comprovação laboratorial para certificar seu diagnóstico. A periculosidade envolvida no atraso do diagnóstico da Dengue demonstra quão importante é uma precoce constatação da infecção no organismo.

O SADDENGUE auxilia o médico no processo de diagnóstico da Dengue. O sistema foi treinado com exemplos reais comprovados por laboratório. O foco do sistema é classificar novos casos de Dengue de acordo com a classificação fornecida pelo médico. Após analisar as características apresentadas pela nova entrada, o sistema retornará a sua classificação.

Vale ressaltar que a acurácia do sistema está intimamente ligada a quantidade de exemplos fornecidos ao seu treinamento, por isso, para trabalhos futuros, sugere-se que sejam adicionados mais dados à base de treinamento para a realização de novos testes e o alcance de uma melhor performance. Além disso, o software necessita de ampliação em sua funcionalidade e aspectos de segurança.

REFERÊNCIAS

ABE, Adriana Helena M.; MARQUES, Solomar M.; COSTA, Paulo Sérgio S. **Dengue em crianças: da notificação ao óbito.** Revista Paulista Pediátrica 30.2 (2012): 263-271.

ALVES, Franciele AS.; NEUMANN, Alexandre MM.; GOUVÊA JR, Maury M. **Bengala Inteligente Neural Baseada em Aprendizagem por Reforço para Deficientes Visuais.** Encontro Nacional de Inteligência Artificial e Computacional, São Carlos (2014).

ANDRADE, PJN de. **Sistemas especialistas de apoio ao diagnóstico em medicina. Relações com o teorema de Bayes e com a lógica do raciocínio diagnóstico.** Arq Bras Cardiol 73.6 (1999): 537-544. Disponível em: <<http://publicacoes.cardiol.br/abc/1999/7306/73060008.pdf>>. Acesso em: 31 mar. 2016.

ANES, Luciana Freitas; FORTES, Reinaldo Silva. **Sistema Especialista para o auxílio no Diagnóstico da Policitemia Vera.** (2013). Disponível em: <<http://www.barbacena.unipac.br/site/bb/tcc/tcc-3fe37ad88c6d85e8193143f6e7d7ef3a.pdf>>. Acesso em: 31 mar. 2016.

BARRETO, Jorge M. **Introdução as redes neurais artificiais.** V Escola Regional de Informática. Sociedade Brasileira de Computação, Regional Sul, Santa Maria, Florianópolis, Maringá (2002): 5-10.

BARRETO, Mauricio; TEIXEIRA, Maria. **Dengue no Brasil: situação epidemiológica e contribuições para uma agenda de pesquisa.** (2008). Disponível em: <<http://www.periodicos.usp.br/index.php/eav/article/view/10348>>. Acesso em: 26 abr. 2016.

BELL, Jason. **Machine Learning: Hands-On for Developers and Technical Professionals.** John Wiley & Sons, Inc. 2015.

BENCH-CAPON, Trevor JM. **Knowledge-based systems and legal applications.** Vol. 36. Academic Press, 2015.

BITTENCOURT, Guilherme. **Inteligência artificial: ferramentas e teorias.** 3. ed. Florianópolis: UFSC, 2006.

BOGDAN, R. S.; BIKEN, S. **Investigação qualitativa em educação: uma introdução à teoria e aos métodos.** 12. ed. Porto: Porto, 2003.

CACIAN, Natália. **Mosquito *Aedes Aegypti***. (2016). Disponível em: <<http://www1.folha.uol.com.br/cotidiano/2016/04/1764824-brasil-ja-tem-ao-menos-91-mil-casos-provaveis-de-zika-rio-lidera-registros.shtml>>. Acesso em: 28 abr. 2016.

CASTILLO, Enrique.; GUTIERREZ, Jose M.; HADI, Ali S. **Expert systems and probabilistic network models**. Springer Science & Business Media, 2012.

CHAVES, M. *et al.* **Dengue, Chikungunya e Zika: a nova realidade brasileira**. (2015). Disponível em: <http://www.newslab.com.br/newslab/revista_digital/132/artigos/artigo1.pdf>. Acesso em: 28 abr. 2016.

CICHOSZ, Pawet. **Data Mining Algorithms: Explained Using R**. John Wiley & Sons, 2015.

CONCEIÇÃO, Elenice.; FARIAS, Edquel.; SASSI, Renato. **Aplicação de técnicas de inteligência artificial no auxílio ao diagnóstico do risco de desenvolver câncer colorretal**. (2014). Disponível em: <<http://repositorio.uninove.br/xmlui/handle/123456789/1010>>. Acesso em: 01 maio 2016.

DAVID, Jean-Marc.; KRIVINE, Jean-Paul.; SIMMONS, Reid. **Second generation expert systems**. Springer Science & Business Media, 2012.

DEAN, Jared. **Big Data, Data Mining And Machine Learning**. John Wiley & Sons, Inc., 2014.

FERREIRA, Tiago Geraldo. **NICeSim: um simulador interativo baseado em técnicas de aprendizado de máquina para avaliação de recém-nascidos prematuros em UTI neonatal**. 2014. Disponível em: <<http://www.locus.ufv.br/bitstream/handle/123456789/2673/texto%20completo.pdf?sequence=1&isAllowed=y>>. Acesso em: 29 abril 2016.

FIOCRUZ, (2016). **Dengue: vírus e vetor**. Disponível em: <<http://www.ioc.fiocruz.br/dengue/textos/sobreovirus.html>>. Acesso em: 26 abr. 2016.

GIL, Antônio Carlos. **Como Elaborar Projetos de Pesquisa**. 4. ed. São Paulo: Atlas, 2006.

GONÇALVES, André Ricardo. **Máquina de Vetores Suporte**. 2015. Disponível em: <<http://www-users.cs.umn.edu/~andre/arquivos/pdfs/svm.pdf>>. Acesso em: 22 maio de 2016.

HALL Mark, Frank Eibe, Holmes Geoffrey, Pfahringer Bernhard, Reutemann Peter, Witten Ian H. (2009); **The WEKA Data Mining Software: An Update**; SIGKDD Explorations, Volume 11, Issue 1.

HAN, Jiawei.; KAMBER, Micheline.; PEI, Jian. **Data mining: concepts and techniques**. Elsevier, 2011.

HOFFMAN, Robert R. **The psychology of expertise: Cognitive research and empirical AI**. Psychology Press, 2014.

HUMMEL, Anderson. *et al.* (2011). **Experiências internacionais da aplicação de sistemas de apoio à decisão clínica em gastroenterologia**. Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/129>>. Acesso em: 01 maio 2016.

KRUSE, Rudolf.; SCHWECKE, Erhard.; HEINSOHN, Jochen. **Uncertainty and vagueness in knowledge based systems: numerical methods**. Springer Science & Business Media, 2012.

LAKATOS, Eva Maria; MARCONI, Marina de Andrade. **Fundamentos de Metodologia Científica**. 7 ed. São Paulo: Atlas, 2010.

LAROSE, Daniel T. **Discovering knowledge in data: an introduction to data mining**. John Wiley & Sons, 2014.

LEWIS, Frank L.; LIU, Derong. **Reinforcement learning and approximate dynamic programming for feedback control**. Vol. 17. John Wiley & Sons, 2013.

LUZ, Kleber *et al.* **Febre pelo vírus Zika**. Epidemiologia e Serviços de Saúde, v. 24, n. 4, p. 785-788. (2015). Disponível em: <http://www.scielo.org/article_plus.php?pid=S2237-96222015000400785&tlng=pt&lng=en>. Acesso em: 28 abr. 2016.

MARCUS, Sandra. **Automating knowledge acquisition for expert systems**. Vol. 57. Springer Science & Business Media, 2013.

MICHALSKI, Ryszard S.; CARBONELL, Jaime G.; MITCHELL, Tom M. **Machine learning: An artificial intelligence approach**. Springer Science & Business Media, 2013.

MINISTÉRIO DA SAÚDE. **Programa mais médicos – dois anos: mais saúde para os brasileiros**. Brasília: Ministério da Saúde, 2015. Disponível em: <http://cdnmaismedicos.elivepress.com.br/images/PDF/Livro_2_Anos_Mais_Medicos_Ministerio_da_Saude_2015.pdf>. Acesso em: 31 mar. 2016.

MINISTÉRIO DA SAÚDE. **Orientações gerais prevenção e combate: informações técnicas**. Brasília: Ministério da Saúde, 2016. Disponível em: <<http://portalsaude.saude.gov.br/index.php/informacoes-tecnicas-dengue>>. Acesso em: 20 set. 2016.

MLAKAR, Jernej *et al.* **Zika virus associated with microcephaly**. New England Journal of Medicine, v. 374, n. 10, p. 951-958. (2016). Disponível em: <<http://www.nejm.org/doi/full/10.1056/NEJMoa1600651>>. Acesso em: 28 abr. 2016.

MOREIRA, Leonardo B.; NAMEN, Anderson A. **Sistema preditivo para a doença de Alzheimer na triagem clínica**. J. Health Inform. (2016) Julho-Setembro; 8(3):87-94. Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/viewFile/413/269>>. Acesso em: 21 set. 2016.

MOURATO, Felipe A.; MOSER, Lúcia R.; MATTOS, Sandra. **Sistema interativo em ambiente móvel para o diagnóstico diferencial de cardiopatias congênitas**. Journal of Health Informatics, v. 5, n. 4, (2013). Disponível em: <<http://www.jhi-sbis.saude.ws/ojs-jhi/index.php/jhi-sbis/article/view/278>>. Acesso em: 21 set. 2016.

MURPHY, Kevin P. **Machine learning: a probabilistic perspective**. MIT press, 2012.

OLIVEIRA JR, Hime A., *et al.* **Inteligência Computacional aplicada a administração, economia e engenharia em Matlab**. Rio de Janeiro: Thompson, 2007.

ROKACH, Lior; MAIMON Oded. **Data mining with decision trees: theory and applications**. World scientific, 2014.

SABBATINI, R. M. **Uso do computador no apoio ao diagnóstico médico**. Revista Informédica, 1(1), 5-11. (1993). Disponível em: <<http://www.informaticamedica.org.br/informed/decisao.htm>>. Acesso em: 31 mar. 2016.

SÁ-SILVA, Jackson Ronie; ALMEIDA Cristóvão Domingos de; GUINDANI, Joel Felipe. **Pesquisa documental: pistas teóricas e metodológicas**. Revista Brasileira de História & Ciências Sociais, 2015. Disponível em: <<http://www.rbhcs.com/rbhcs/article/view/6/pdf>>. Acesso em: 15 maio 2016.

SECRETARIA DE VIGILÂNCIA EM SAÚDE - MINISTÉRIO DA SAÚDE. **Febre pelo vírus Zika: uma revisão narrativa sobre doença**. V.46, n.26, p.4 (2015). Disponível em: <<http://portalsaude.saude.gov.br/images/pdf/2015/agosto/26/2015-020-publica----o.pdf>>. Acesso em 28 abr. 2016.

SINGHI, Sunit; KISSOON, Niranjan; BANSAL, Arun. **Dengue e dengue hemorrágico: aspectos do manejo na unidade de terapia intensiva**. Jornal de Pediatria, v. 83, n. 2, p. S22-S35, 2007.

SOL, H. G.; CEES, A. T.; DE VRIES ROBBÉ, P. F. (2013). **Expert systems and artificial intelligence in decision support systems**: proceedings of the Second Mini Euroconference, Lunteren, The Netherlands, 17–20 November 1985. Springer Science & Business Media.

SOMMERVILLE, Ian. **Engenharia de Software**, 8ª edição, Tradução: Selma Shin Shimizu Mel-nikoff, Reginaldo Arakaki, Edilson de Andrade Barbosa. São Paulo: Pearson Addison-Wesley, 2007.

SRIVASTAVA, Shweta. **Weka: a tool for data preprocessing, classification, ensemble, clustering and association rule mining**. International Journal of Computer Applications 88.10, 2014.

TEIXEIRA, Maria, *et al.* **Dengue: Vinte e Cinco anos da reemergência no Brasil**. (2009). Cadernos de saúde pública, vol. 25 suppl.1 Rio de Janeiro. Disponível em: <http://www.scielo.org/scielo.php?script=sci_arttext&pid=S0102-311X2009001300002>. Acesso em: 26 abr. 2016.

WECHSLER, Rudolf, *et al.* **A informática no consultório médico**. J Pediatr (Rio J) 79.Supl 1 (2003): 3-12. Disponível em: <<http://www.scielo.br/pdf/%0D/jped/v79s1/v79s1a02.pdf>>. Acesso em: 31 mar. 2016.

WORLD HEALTH ORGANIZATION (2011). **Comprehensive Guidelines for Prevention and Control of Dengue and Dengue Hemorrhagic Fever**. Regional Office for South-East Asia. (Índia) SEARO Technical Publication Series No. 60. Disponível em: <http://apps.searo.who.int/pds_docs/B4751.pdf>. Acesso em: 26 abr. 2016.

ANEXO A – DOCUMENTO DE SOLICITAÇÃO / FRENTE

**MINISTÉRIO DA EDUCAÇÃO
SECRETARIA DE EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA
INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE MINAS GERAIS
CAMPUS SÃO JOÃO EVANGELISTA**

Avenida Primeiro de Junho, nº 1043 – Bairro Centro – São João Evangelista – Minas Gerais – CEP: 39.705-000
(33) 3412-2900

Ofício nº. 0010/2016 – DDE/CGEMT/CAMPUS-SJE/IFMG/SETEC/MEC

São João Evangelista, 16 de junho de 2016

Ilma. Sra.
RAYLAINE CASTRO
Gerente de Vigilância Epidemiológica de Governador Valadares- MG

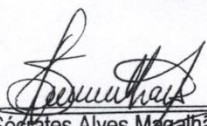
Assunto: **Solicitação que se faz.**

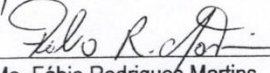
Prezada Ilma. Sra.

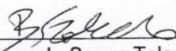
Vimos, muito respeitosamente, solicitar deste setor de Vigilância Epidemiológica, de Governador Valadares-MG, dados de pacientes diagnosticados com as doenças: *Dengue, Zika e Chikungunya*, para serem usados como base de treinamento de um sistema de apoio ao diagnóstico médico das enfermidades supracitadas. Esse estudo está sendo desenvolvido como Trabalho de Conclusão de Curso (TCC) que é um requisito para obtenção do título de Bacharelado em Sistemas de Informação pelo Instituto Federal de Minas Gerais, Campus São João Evangelista (IFMG-SJE). O mesmo está sendo desenvolvido pela equipe de 4 pessoas sendo dois discentes, Sócrates Alves Magalhães CPF: 099.029.946-52 e Vânia Batista dos Santos CPF: 018.132.766-09, e dois docentes, orientador Ms. Fábio Rodrigues Martins e coorientador Dr. Wesley Gomes de Almeida. O estudo encontra-se em andamento e depende no momento desses dados para sua efetivação. Dados esses, referentes aos sintomas apresentados pelos diagnosticados, assim como as características que não identificam o paciente, mas que são de suma importância para o conjunto de dados, tais como: idade, sexo, os sintomas apresentados, e a confirmação de cada caso com positivo ou negativo para as três doenças correlatas.

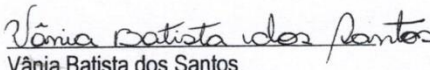
Sendo o que temos para o momento, agradecemos e nos colocamos à disposição para quaisquer esclarecimentos que por ventura sejam necessários.

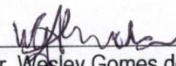
ANEXO B – DOCUMENTO DE SOLICITAÇÃO / VERSO

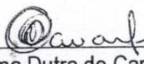

 Sócrates Alves Magalhães
 Bacharelado em SI IFMG-SJE.


 Me. Fábio Rodrigues Martins
 Orientador do projeto IFMG-SJE


 Me. Bruno de Souza Toledo
 Coordenador do Curso de Bacharel
 em Sistemas de Informação IFMG-SJE


 Vânia Batista dos Santos
 Bacharelada em SI IFMG-SJE.


 Dr. Wesley Gomes de Almeida
 Coorientador do projeto IFMG-SJE


 Ma. Karina Dutra de Carvalho Lemos
 Diretora do Departamento de Ensino do
 IFMG-SJE

Karina D. de Carvalho Lemos
 Diretora Depto. de Ensino
 Port. IFMG-SJE 042/2015

Recbi em 17/06/16


 Raylaine Castro dos Santos
 Coordenadora Interina Epidemiológica
 Departamento de Vigilância em Saúde
 GEPI/DVS-GV