FACULDADE DE TECNOLOGIA DE AMERICANA

Curso Superior de Tecnologia em Segurança da Informação

Bruna Luisa do Amaral da Silva

REVISÃO SISTEMÁTICA SOBRE DEEP LEARNING APLICADO A DETECÇÃO DE CYBER ATAQUES

Americana, SP

2018

FACULDADE DE TECNOLOGIA DE AMERICANA

Curso Superior de Tecnologia em Segurança da Informação

Bruna Luisa do Amaral da Silva

REVISÃO SISTEMÁTICA SOBRE DEEP LEARNING APLICADO A DETECÇÃO DE CYBER ATAQUES

Trabalho de Conclusão de Curso desenvolvido em cumprimento à exigência curricular do Curso Superior de Tecnologia em Segurança da Informação, sob a orientação do Dr. Kleber de Oliveira Andrade

Área de concentração: Segurança da Informação

Americana, SP

2018

FICHA CATALOGRÁFICA – Biblioteca Fatec Americana - CEETEPS
Dados Internacionais de Catalogação-na-fonte

S578r SILVA, Bruna Luisa do Amaral da

Revisão sistemática sobre deep learning aplicado a detecção de cyber ataques. / Bruna Luisa do Amaral da Silva. – Americana, 2018.

44f.

Monografia (Curso de Tecnologia em Segurança da Informação) - - Faculdade de Tecnologia de Americana – Centro Estadual de Educação Tecnológica Paula Souza

Orientador: Prof. Dr. Kleber de Oliveira Andrade

1 Segurnaça em sistemas de informação I. ANDRADE, Kleber de Oliveira II. Centro Estadual de Educação Tecnológica Paula Souza – Faculdade de Tecnologia de Americana

CDU: 681.518.5

Bruna Luisa do Amaral da Silva

REVISÃO SISTEMÁTICA SOBRE DEEP LEARNING APLICADO A DETECÇÃO DE CYBER ATAQUES

Trabalho de graduação apresentado como exigência parcial para obtenção do título de Tecnólogo em Segurança da Informação pelo CEETEPS/Faculdade de Tecnologia – FATEC/ Americana.
Área de concentração: Segurança da Informação

Americana, 28 de junho de 2018.

Banca Examinadora:

Kleber de Oliveira Andrade (Presidente)
Doutor
FATEC Americana

Henri Alves de Godoy (Membro)
Mestre
FATEC Americana

Renato Kraide Soffner (Membro)
Doutor
FATEC Americana

AGRADECIMENTOS

Em primeiro lugar gostaria de agradecer à minha família e amigos, que sempre acreditaram na minha capacidade e me apoiam em todos os momentos.

Gostaria de agradecer a toda equipe administrativa da faculdade e a todos os professores, em especial ao meu orientador, que me ensinou e me deu todo o suporte necessário para que este trabalho fosse realizado.

DEDICATÓRIA

Aos meus pais, ao meu irmão e ao meu marido.

RESUMO

O presente trabalho tem como objetivo situar o leitor sobre a situação atual das pesquisas relacionadas a aplicação de *deep learning* para detecção de *cyber* ataques para tal, inicialmente é explicado conceitos básicos referente a inteligência artificial, aprendizado de máquina, *deep learning* e *cyber* ataques, além de descrever os processos envolvidos na Revisão Sistemática de Literatura. Estudos relacionados ao tema foram selecionados, e então foi feita uma avaliação para identificar quais eram relevantes ou não para a revisão. Com a análise dos estudos aceitos, é possível identificar através da métrica de acuracidade a efetividade da aplicação de algortimos de *deep learning* na detecção dos mais diversos tipos de *cyber* ataques como, *malwares*, ataques de intrusão, ataques *web*, detecção de anomalias e ataques em redes sem fio. O resultado médio alcançado pelos trabalhos analisados foi de 95% de acurácia.

Palavras Chave: *Cyber* Ataque, *Deep Learning*, Revisão Sistemática

ABSTRACT

The present work aims to situate the reader on the current situation of research related to the application of deep learning to detect cyber-attacks for this, initially it explains basic concepts regarding artificial intelligence, machine learning, deep learning and cyber-attacks, in addition the process of systematic review is described. Studies related to the topic were selected, and then an evaluation was made to identify which were relevant or not for the review. With the analysis of the accepted studies, it is possible to identify through the accuracy metric the effectiveness of the application of deep learning algorithms in detection of the most diverse types of cyber-attacks such as malware, intrusion attacks, web attacks, anomaly detection and attacks on wireless networks. The average result for the analyzed works was 95% of accuracy.

Keywords: *Cyber-Attacks; Deep Learning; Systematic Review*

SUMÁRIO

LISTA DE FIGURAS

LISTA DE TABELAS

1 INTRODUÇÃO

A inteligência artificial tem sido tendência em diversos setores de mercado e áreas de estudo, e para segurança da informação não é diferente, de acordo com pesquisas da Gartner (2017) o uso de inteligência artificial para segurança da informação é mais que uma tendência e sim uma necessidade.

“As mudanças em segurança digital vão requerer novos tipos de habilidades em *Data Science* e *Analytics*. O crescimento em informação significa que a Inteligência Artificial é necessária.”

A IA é definida basicamente como uma máquina com capacidade de realizar atividades que até então apenas humanos poderiam realizar. Russell e Norvig (2012), descrevem a Inteligência Artificial em quatro aspectos: pensar como humano, agir como humano, pensar racionalmente, e agir racionalmente.

Outra tecnologia em ascensão, é um subcampo da IA, o Aprendizado de Máquina, que de acordo com Segaran (2007), é um campo de estudos o qual se preocupa em estudar e criar algoritmos que permitam computadores aprender. Neste subcampo, podemos ainda encontrar uma outra área que vem ganhando cada vez mais destaque, o *Deep Learning* (Aprendizado Profundo, em tradução literal), o qual possibilita que a máquina aprenda conceitos complexos, os construindo a partir de conceitos mais básicos (GOODFELLOW et al, 2016).

Se tratando de Segurança da Informação em específico, os *cyber* ataques, tem tirado o sono e o dinheiro de muita gente, de acordo com o CSO Online (2018), é esperado que o *cyber* crime atinja um prejuízo de US$ 6 trilhões por ano até 2021.

Com todos estes holofotes voltados para estas áreas de pesquisa e toda a preocupação com os prejuízos que podem ser causados pelos *cyber* ataques, muitos trabalhos têm sido realizados com enfoque na utilização de DL aplicado à SI. Ao realizar uma pesquisa rápida na página do Google Acadêmico pelos termos “*Deep Learning for Cyber Security”* com datas a partir de 2010 até maio de 2018, o resultado bruto apresentado é de mais de 17 mil estudos.

A quantidade de trabalhos é realmente impressionante, porém isto não significa necessariamente que avanços reais estão sendo alcançados com todas estas pesquisas. Com isso surge a seguinte questão: Qual é a efetividade da utilização de algoritmos de *deep learning* para a detecção de *cyber* ataques?

E para responder a esta pergunta foi percebida a necessidade de realizar uma revisão sistemática, a fim de sumarizar as evidências existentes em trabalhos primários, analisar seus resultados e destacar principais problemas e soluções por eles apresentados.

O trabalho foi estruturado em cinco capítulos, sendo primeiro reservado a esta introdução, o segundo conceitua a revisão sistemática, mostrando através de um passo a passo, como a mesma funciona e os critérios utilizados para a seleção dos trabalhos em análise, no terceiro capítulo temos a descrição de alguns conceitos básicos de IA, AM, DL e *cyber* ataques, no quarto capítulo temos um resumo dos trabalhos que são analisados e podemos identificar os resultados apresentados, e com base nas informações conseguidas a partir das pesquisas realizadas para os capítulos anteriores, o capítulo cinco se reserva às considerações finais.

2 REVISÃO SISTEMÁTICA

A revisão de literatura sistemática é uma forma de estudo secundário, que utiliza de uma metodologia bem definida para identificar, analisar e interpretar todas as evidências disponíveis relacionadas à uma pergunta de pesquisa específica. (KITCHENHAM E CHARTERS, 2007).

Ainda de acordo com Kitchenham e Charters (2007), algumas vantagens ao se aplicar a revisão sistemática são:

i. Por ser uma metodologia em que todos os critérios são muito bem definidos, a resuz a possibilidade de que a estudo seja enviesado.
ii. Pode-se identificar através da revisão sistemática, que se um estudo obtiver resultados consistentes, o mesmo é robusto e transferível, porém se o estudo apresenta resultados inconsistentes, podem ser analisadas as fontes de varição.

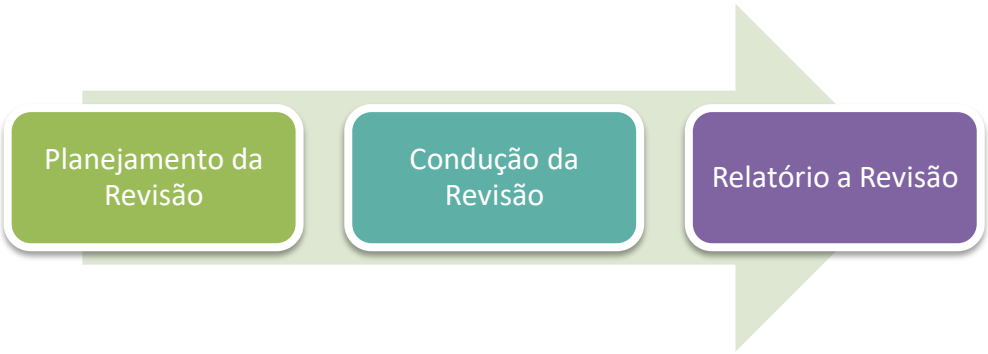
Este é um tipo de estudo secundário, o qual se baseia em uma ampla base de pesquisas primárias existente, por esta razão é muito importante explicitar todos os critérios de inclusão e de exclusão, assim pode-se determinar a qualidade dos estudos utilizados como fonte para a revisão.

2.1 Etapas da revisão sistemática

Existem três etapas básicas na construção da revisão sistemática, que são realizados em modo sequencial, porém havendo a possibilidade, e em alguns casos a necessidade, de que sejam realizados em modo paralelo (MUNZLINGER, NARCIZO E QUEIROZ, 2012).

Para cada item das etapas descritas na Figura 1, existem alguns subitens que darão o direcionamento para a condução da revisão sistemática.

Figura 1 – Etapas da revisão sistemática



Fonte: Adaptada de Kitchenham e Charters (2007).

O primeiro passo é o planejamento da revisão, onde é identificada a necessidade do estudo, qual será a pergunta respondida por ele e quais serão os critérios aplicados na pesquisa. No segundo passo temos a condução da revisão, este é o momento de execução da pesquisa, com seleção de material, a avaliação da qualidade dos estudos, extração e monitoramento dos dados e a síntese dos mesmos. Por fim, no terceiro e último passo, deve-se sumarizar os resultados e suas interpretações através de relatórios.

2.1.1 Planejamento da revisão

A primeira etapa do processo de revisão sistemática, é o planejamento da revisão (Tabela 1), que se inicia pela identificação da necessidade da mesma, que deverá especificar o porquê essa pesquisa faz-se necessária, qual é o avanço que este estudo pode trazer para a comunidade científica.

Tabela 1 – Planejamento da revisão (etapa 1)

Planejamento da Revisão	Identificação da necessidade de uma revisão
	Especificar a pergunta de pesquisa
	Desenvolver o protocolo da revisão

Fonte: Tabela adaptada de Kitchenham e Charters (2007).

Então podemos definir a pergunta de pesquisa, que é uma das partes de maior importância em uma revisão sistemática, a pergunta direciona todo o estudo que será realizado.

Logo após a definição da pergunta, deverá ser criado o protocolo da revisão, que conta com:

i. A justificativa de pesquisa.
ii. A pergunta da pesquisa.
iii. A estratégia que será usada para seleção de estudos primários, incluindo termos e métodos de pesquisa.
iv. Critérios de seleção do estudo. Os critérios de seleção do estudo são usados para determinar quais estudos estão incluídos ou excluídos de uma revisão sistemática. Exemplos de critérios: idioma, data de publicação, autores, método de pesquisa, entre outros.
v. Procedimentos de seleção de estudo. O protocolo deve descrever como os critérios de seleção serão aplicados.
vi. Lista de avaliação de qualidade dos estudos e procedimentos.
vii. Estratégia de extração de dados. Define como a informação requerida de cada estudo primário será obtida.
viii. Síntese dos dados extraídos. Define a estratégia de síntese

2.1.2 Condução da revisão

Na segunda etapa (Tabela 2), inicia o processo de revisão, com a identificação e seleção de pesquisa, que tem como objetivo reunir a maior quantidade possível de estudos utilizando uma estratégia de pesquisa imparcial. O rigor no processo de pesquisa, é um dos fatores que diferencia a revisão sistemática de métodos tradicionais de revisão. É importante realizar a documentação destas pesquisas para que seja um processo transparente e replicável.

Outro fator crítico para a revisão sistemática, é a avaliação de qualidade das pesquisas, e apesar de não haver um consenso na definição de qualidade dos estudos, Cochrane Reviewers' Handbook (2004), sugere que a qualidade está relacionada à minimizar a parcialidade e maximizar a validade interna e externa.

Os formulários de extração de dados devem ser projetados para coletar todas as informações necessárias para abordar as questões de revisão e os critérios de qualidade do estudo.

Na síntese dos dados, as informações referentes aos estudos devem ser apresentadas de forma a evidenciar as similaridades e as diferenças entre os resultados apresentados pelos estudos.

Tabela 2 – Condução da revisão (etapa 2)

Condução da Revisão	Identificação das pesquisas
	Seleção de pesquisas primárias
	Avaliação de qualidade das pesquisas
	Extração e monitoramento de dados
	Síntese de dados

Fonte: Tabela adaptada de Kitchenham e Charters (2007).

2.1.3 Relatório da revisão

Por fim, os trabalhos inclusos na revisão sistemática podem ser apresentados em um quadro que destaca suas características principais, tais como, autores, ano de publicação, metodologia, variáveis dependes e principais resultados. A Tabela 3 apresenta um resumo.

Tabela 3 – Relatório da revisão (etapa 3)

Relatório da Revisão	Formatar relatório principal
	Avaliar o relatório

Fonte: Tabela adaptada de Kitchenham e Charters (2007).

3 CONCEITOS FUNDAMENTAIS

Este capítulo expõe os conceitos fundamentais relacionados a este trabalho. Serão apresentadas as definições de Inteligência Artificial, Aprendizado de Máquina, *Deep Learning* e *Cyber* Ataques.

3.1 Inteligência Artificial

Russell e Norvig (2012) descrevem com base em citações próprias e de outros autores, a definição de IA em quatro aspectos: pensar como humano, agir como humano, pensar racionalmente, e agir racionalmente.

i. **Agir como humano:** De acordo com o teste de Turing, proposto por Alan Turing (1950), um computador passa no teste se um interrogador humano, não puder afirmar se as respostas vieram de um computador ou de uma pessoa. Para que um computador consiga passar pelo teste, ele precisará possuir as seguintes capacidades: Processamento de Linguagem Natural, Representação de Conhecimento, Raciocínio Automatizado, Aprendizado de Máquina, Visão Computacional e Robótica. Estas seis disciplinas compõe a maior parte da Inteligência Artificial.

ii. **Pensar como humano:** Ainda de acordo com Russell e Norvig (2012), se quisermos que programas pensem como humanos, nós devemos ter formas de determinar como os humanos pensam. Nós precisamos estar por dentro dos trabalhos atuais relacionados a mente humana. Existem três formas de fazermos isto: através da introspecção – tentar capturar nossos próprios pensamentos conforme eles vão surgindo; através de experimentos psicológicos – observando uma pessoa em ação; e por imagem cerebral – observando (através de ressonância magnética, por exemplo) o cérebro em ação.

iii. **Pensar racionalmente:** Baseado no silogismo (raciocínio dedutivo) de Aristóteles, foi iniciado o estudo do campo da lógica. Russell e Norvig (2012) mencionam que lógicos do século XIX desenvolveram notações de afirmações precisas sobre todos os tipos de objetos no mundo e as

relações entre eles. Em 1965, existiam programas que podiam resolver qualquer problema solucionável descrito em notação lógica. (Porém se a solução não existisse, o programa possivelmente entraria em um loop infinito). A então chamada tradição logicista (progressão do pensamento que se dá entre premissas) dentro da inteligência artificial espera construir programas para criar sistemas inteligentes.

iv. **Agir Racionalmente:** Um agente é apenas algo que age (agente vem do Latim agere, fazer). Claro que, computadores fazem algo, mas é esperado mais de computadores agentes, como, operar de forma autônoma, perceber seu ambiente, durar por um longo período de tempo, se adaptar a mudanças, e criar e perseguir objetivos. Um agente lógico é aquele que trabalha para atingir o melhor resultado, ou, quando houver incerteza, o melhor resultado possível.

Segundo Russell e Norvig (2012), a IA é um campo de estudo no qual se relacionam os mais variados campos de conhecimento, como: Filosofia, Matemática, Economia, Neurociência, Psicologia, Engenharia da Computação, Teoria de Controle e Cibernética e Linguística.

De modo geral dentro de cada uma dessas áreas de conhecimento são realizadas algumas perguntas que devem ser respondidas para que haja um entendimento do funcionamento do cérebro humano e assim tornar possível reproduzir ou até mesmo melhorar as ações e pensamentos humanos em máquinas.

3.2 Aprendizagem de máquinas

O Aprendizado de Máquina é um subcampo da IA, que de acordo com Segaran (2007), se preocupa em estudar e criar algoritmos que permitam computadores aprender. Isso significa que, na maioria dos casos, este é um algoritmo ao qual é informado um conjunto de dados e informações sobre a propriedade dos dados, e essa informação permite fazer predições sobre outros dados que podem ser acessados no futuro. Isto é possível, pois a maioria dos dados, contém padrões, e esses padrões permitem que a máquina generalize. A fim de generalizar, a máquina treina um modelo com o qual determina aspectos importantes dos dados.

3.3 *Deep Learning*

Goodfellow et al (2016), diz que um dos maiores desafios para o campo da Inteligência Artificial, é o de resolver problemas que para os humanos é uma tarefa simples de ser realizada, porém difícil de ser explicada, pois são feitas de forma intuitiva, quase automaticamente.

Portanto é desenvolvida uma abordagem que possibilita a máquina a aprender a partir das experiências e entendimento do mundo em termos de conceitos de hierarquia, assim o computador aprende conceitos complexos construídos com base em conceitos mais simples. O grafo gerado deste método, é profundo, com muitas camadas, e por este motivo é chamado de Aprendizado Profundo, do inglês, *Deep Learning* que por sua vez é um subcampo do Aprendizado de Máquina.

Assim como o Aprendizado de Máquina, o *Deep Learning* possui seus vários algoritmos os quais tem suas vantagens e desvantagens para os mais variados tipos de aplicação.

Os estudos revisados neste trabalho, abordam os algoritmos de DL descritos abaixo.

i. ***Convolutional Neural Network***: as redes neurais convolucionais, é um *perceptron* (um classificador binário que mapeia a entrada para um valor de saída através de uma matriz) multicamada, que tem como objetivo reconhecer formas bidimensionais com alto grau de invariância à tradução, escala, inclinação e outras formas de distorção (HAYKIN, 2009).

ii. ***Deep Belief Networks***: de acordo com Deng e Yu (2014) DBNs são modelos generativos probabilísticos compostos de múltiplas camadas de variáveis estocásticas e ocultas. O topo das duas camadas tem conexões simétricas não direcionadas entre elas. As camadas inferiores recebem conexões direcionadas da camada acima.

iii. ***Recurrent Neural Networks (RNNs)***: são modelos de redes neurais artificiais que são adequadas para classificação padrões em tarefas cujas entradas e saídas são sequências. Uma RNN representa uma sequência com um vetor de alta dimensão (chamado de estado oculto) de um vetor fixo dimensionalidade que incorpora novas observações usando uma função intrincada não-linear (SUSTSKEVER, 2013).

iv. ***Self-taught Learning***: este é um algoritmo de DL que consiste em dois estágios de classificação. Primeiramente, a representação de características é aprendida com base em uma grande coleção de dados não classificados. Na segunda parte do processo, a representação aprendida é aplicada para dados rotulados e usada para tarefas de classificação (JAVAID et al,2016).

v. ***Deep Neural Network***: as redes neurais artificiais foram incialmente apresentadas como uma modelo computacional de como as redes neurais biológicas, deveriam funcionar de maneira que opressões computacionais complexas pudessem ser realizadas utilizando a lógica proposicional (lógica na qual os fatos são apresentados em proposições). Quando uma rede neural artificial possui duas ou mais camadas escondidas, ela é chamada *Deep Neural Network* (GÉRON, 2018).

vi. ***AutoEncoders***: uma rede neural de *autoencoder* é um algoritmo de aprendizado de máquina não supervisionado que aplica a retropropagação, definindo os valores de destino como iguais às entradas. De acordo com Goodfellow et al (2016) um autoencoder é treinado para tentar copiar sua entrada para sua saída. Internamente, ele possui uma camada oculta que descreve um código usado para representar a entrada.

vii. ***H2O Deep Learning Algorithm***: é um algoritmo de *deep learning* baseado em rede neural artificial *feedforward* (estrutura na qual os nós entre as conexões da rede não formam um círculo), que é treinado com gradiente descendente estocástico usando retropropagação. Cada nó de cálculo treina uma cópia dos parâmetros do modelo global em seus dados locais com multi-encadeamento (de forma assíncrona) e contribui periodicamente para o modelo global por meio da média do modelo na rede (H2O Documentation 2018).

3.4 *Cyber* ataques

A abrangência de ameaças em ambiente cibernético é vasta, portanto muitos estudos são realizados tentando solucionar um problema em específico, neste trabalho serão abordados os seguintes tipos de ataques: *Malwares*, Intrusão de Redes e/ou Sistemas, Ataques em Redes sem Fio e Ataques Web.

i. **Malwares:** de acordo com definição do site da Norton (2018), *malware* é a abreviação de *software* malicioso, do inglês, *malicious software*. Sendo este um sistema feito especificamente para ganhar acesso ou causar dano a um computador (ou outro dispositivo) sem o conhecimento do dono. Existem vários tipos de *malwares* incluindo *spyware*, *keyllogers*, vírus, *worms* ou qualquer outro tipo de código malicioso que possa infectar o dispositivo.

ii. **Intrusão de Redes e/ou Sistemas:** uma intrusão é qualquer tipo de atividade não autorizada em uma rede e/ou sistema (RSA 2018).

iii. **Ataques em Redes sem Fio:** este tipo de ataque utiliza de redes sem fio para ganhar acesso ou causar danos a um dispositivo ou rede, os principais tipos de ataques em redes sem fio segundo Uhcôa (2015), são: *scanning, sniffers, spoofing e denial of service.*

iv. **Ataques Web:** De acordo com a OWASP (2017) os principais ataques web atualmente são: *injection, broken authentication, sensitive data exposure, external enteties e broken access control*.

4 RESULTADOS

Para seleção dos estudos em análise neste trabalho, foi utilizada a ferramenta StArt 3.4 (*State of Art through Systematic Review*) da UFSCAR, a qual tem como objetivo, auxiliar pesquisadores a realizar revisões sistemáticas de forma mais rápida e com maior qualidade.

4.1 Protocolo de pesquisa

O primeiro passo para iniciar o processo na ferramenta, foi fazer o preenchimento de um protocolo contendo, a descrição dos objetivos, a questão de pesquisa, critérios de seleção, métodos de busca, tipo do estudo e métodos de avaliação. Sendo eles:

i. **Objetivo:** O objetivo deste trabalho é identificar através da revisão sistemática a eficácia de algoritmos de deep learning na detecção de cyber ataques.
ii. **Questão Principal:** Qual tem sido a efetividade de algoritmos de deep learning na detecção de cyber ataques?
iii. **Palavras-Chave:** *Artificial Intelligence*, *Attack*, *Cyber Attack*, *Cyber Attacks*, *Cyber Security*, *Cybersecurity*, *Deep Learning*, *Information Security*, *Intrusion Detection*, *Malware*, *Deep Neural Network*, *Network Intrusion*, Inteligência Artificial, *Cyber* Ataques.
iv. **Idiomas:** Português ou Inglês.
v. **Método de busca:** Busca através de pesquisas na internet.
vi. **Tipo de Estudo:** Qualitativo.
vii. **Método de Avaliação:** Percentual de Acuracidade.
viii. **Critérios de inclusão/exclusão:** Apresentados na Tabela 4.

Tabela 4 – Critérios de Inclusão e Exclusão para a Revisão Sistemática

Critérios de Inclusão	Critérios de Exclusão
Artigos completos sobre detecção de *cyber* ataques utilizando algoritmos e métodos de *deep learning*.	Artigos completos sobre detecção de *cyber* ataques utilizando outras técnicas.
Artigos com data de publicação entre 2010 e 2018.	Artigos completos com outras datas de publicação.
Artigos nos idiomas português ou inglês.	Artigos em outros idiomas que não sejam português ou inglês.
	Artigos sobre assuntos relacionados à segurança que não sejam detecção de cyber ataques.

Fonte: Elaborado pelo autor

E então a partir do protocolo criado, são feitas as buscas em base de dados confiáveis, utilizando filtros para seleção de estudos que atendam os critérios pré-definidos.

Para a busca foram considerados trabalhos primários, com tema de pesquisa relacionado a detecção de *cyber* ataques utilizando técnicas de *deep learning*, artigos em inglês ou português (apesar de não conseguir coletar trabalhos em português que atendessem todos os critérios) e com datas de publicação a partir de 2010 até 30 de maio de 2018.

As bases de pesquisas das quais os estudos foram retirados foram: Google Acadêmico, ACM (*Associention for Computing Machinery*) e *Science Direct*.

A partir deste ponto todas as publicações foram revisadas, com base no título, palavras-chave e resumo, a fim de classificar os trabalhos como aceitos, rejeitados ou duplicados.

Foram classificados como aceitos um total de 18 estudos, os quais podem ser verificados na Tabela 5.

Após a classificação de todos os trabalhos, foi realizada a análise dos mesmos a fim de identificar pontos que respondam à pergunta principal deste estudo.

Tabela 5 – Estudos aprovados para a revisão sistemática

Título	Referência	Acurácia
Deepsign: Deep learning for automatic malware signature generation and classification	David e Netanyahu (2015)	98,60%
A deep learning approach for network intrusion detection system	Javaid et al. (2016)	98,00%
Deep learning approach for network intrusion detection in software defined networking	Tang et al. (2016)	75,75%
Deep4maldroid: A deep learning framework for android malware detection based on linux kernel system call graphs	Hou et al. (2016)	93,68%
Deep learning for unsupervised insider threat detection in structured cybersecurity data streams	Tuor et al. (2017)	95,53%
Detecting impersonation attack in WiFi networks using deep learning approach	Aminanto e Kim (2016)	97,93%
Malware traffic classification using convolutional neural network for representation learning	Wang et al. (2017)	99,41%
Anomaly-Based Web Attack Detection: A Deep Learning Approach	Liang, Zhao, e Ye (2017)	98,56%
An Intrusion Detection Algorithm of Dynamic Recursive Deep Belief Networks	Tian e Li (2017a)	92,46%
DeepLog: Anomaly Detection and Diagnosis from System Logs Through Deep Learning	Du et al. (2017)	99,00%
Design and Implementation of Intrusion Detection System Using Convolutional Neural Network for DoS Detection	Nguyen et al. (2018)	98,56%
Deep Android Malware Detection	McLaughlin et al. (2017)	88,30%

Droid-Sec: Deep Learning in Android Malware Detection	Yuan et al. (2014)	96,50%
CNN-Webshell: Malicious Web Shell Detection with Convolutional Neural Network	Tian et al. (2017b)	98,60%
Malware Classification Using Deep Learning Methods	Cakir e Dogdu (2018)	96,00%
Deep Neural Networks for Automatic Android Malware Detection	Hou et al. (2017)	96,40%
Early-Stage Malware Prediction Using Recurrent Neural Networks	Rhode, Burnap e Jones (2018)	94,00%
Security Analytics: Using Deep Learning to Detect Cyber Attacks	Lambert (2017)	97,11%

Fonte: Elaborado pelo autor

4.2 Apresentação dos resultados

Com os resultados obtidos na revisão sistemática, foi possível identificar algoritmos/técnicas de *deep learning* que tem mostrado ótimos resultados na detecção de cyber ataques.

A métrica utilizada para definir a qualidade das soluções de *deep learning* propostas, foi a acurária. Esta métrica basicamente, diz o quanto o método em geral está correto.

$$Acurácia = \frac{Verdadeiros\ Positivos + Verdadeiros\ Negativos}{Total}$$

4.2.1 Resultados para detecção de *malwares*

No total foram 8 estudos selecionados, nos quais pretende-se solucionar os problemas na detecção de *malwares*, sendo 4 deles específicos para detecção de *malwares* em dispositivos móveis. Neste tópico será descrito de forma resumida as soluções propostas pelos autores e quais foram os resultados alcançados.

No artigo de David e Netanyahu (2015), é a apresentada uma solução chamada *Deepsign* para geração de assinaturas e classificação de *malwares*, baseada em *Deep Belief Network* e *AutoEncoders*, assim tornando possível gerar uma representação invariante do comportamento do *malware*. A base de dados utilizada para treinamento desta solução de *deep learning*, continha seis categorias de *malwares* (Zeus, *Carberp, SpyEye, Cidox*, Andromeda, e *DarkComet*), com um total de 1.800 exemplos. A solução se mostrou bastante efetiva, conseguindo atingir uma acurácia de 98.6%.

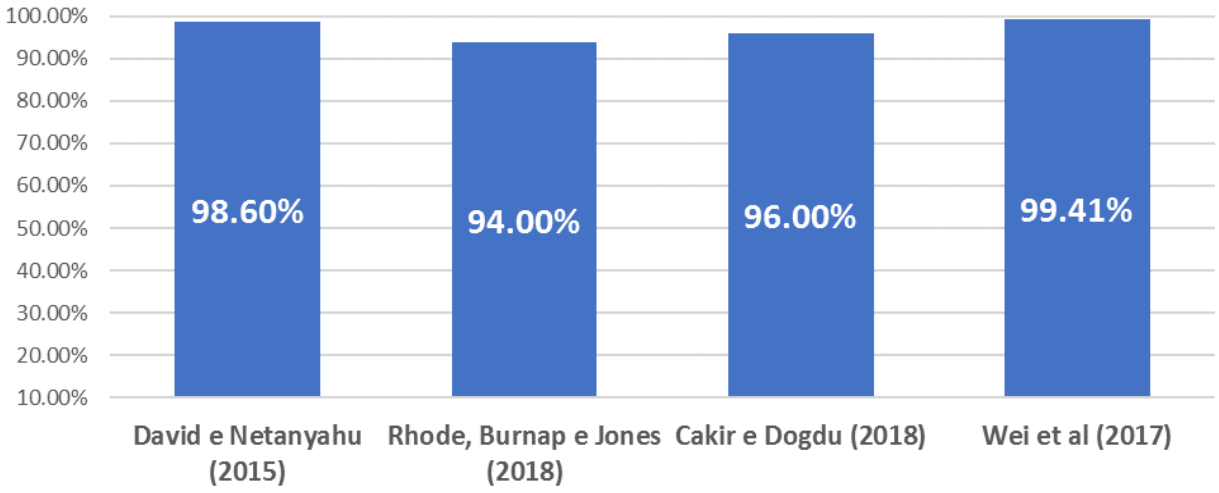
Outro estudo referente a detecção de *malware*, corrobora com o trabalho apresentado acima, pois também alcança bons resultados utilizando *deep learning*, foi realizado por Rhode, Burnap e Jones (2018), com o algoritmo *Recurrent Neural Network*, o qual demonstrou a capacidade de identificar se um executável era malicioso ou não com tempo médio de 5 segundos. Os pesquisadores utilizaram um conjunto de dados, contendo 2.286 exemplos maliciosos e 2.345 não maliciosos. Este modelo atingiu uma acurácia de 94%.

No terceiro estudo analisado, os autores Cakir e Dogdu (2018), deixam claro um “problema” na utilização dos métodos de *deep learning*, os mesmos necessitam de mais tempo computacional para treino e retreino dos modelos, enquanto os algoritmos de aprendizado de máquina levam bem menos tempo, em contrapartida apesar de algoritmos de AM precisarem de menos tempo, eles acabam alcançando uma acurácia menor. Entretanto é enfatizado que houve um bom desempenho do método utilizado, que é baseado em *Convolutional Neural Network*, *AutoEncoder* e *Recurrent Neural Network,* que atingiu entre 94% e 96% de acuracidade. A base de dados utilizada para este estudo foi a *Microsoft Malware Classification Challenge Dataset*.

Wang et al (2017) propõe um método de classificação de tráfego de *malware* aplicando *Convolutional Neural Network*, o qual utiliza imagens dos dados crus do tráfego da rede como entrada para o classificador. Este estudo foi o que obteve o melhor resultado, o qual ficou com 99,41% de acuracidade. A base de dados utilizada, foi criada pelos próprios autores USTC-TFC2016 - que foi dívidida em duas partes:

i. Parte 1: Dados reais de tráfego malicioso em website público coletado de CTU (https://www.cvut.cz/en/science-and-research-at-ctu) entre 2011 e 2015;
ii. Parte 2: Tráfego normal coletado através de IXIA BPS.

Figura 2 – Acurácia para Detecção de Malware



Fonte: Elaborado pelo autor

Os próximos quatro artigos que serão abordados, referem-se a detecção de *malware* em dispositivos móveis.

McLaughlin et al (2017) apresentam um sistema de detecção de malware para dispositivos *android* baseado em *Convolutional Neural Network*, sendo destacado pelos autores que um dos diferenciais do sistema por eles apresentados é que ele elimina a necessidade de criação manual de recursos de *malware*. Para este estudo foi utilizada uma base independente de dados com exemplos de aplicativos maliciosos e não maliciosos. Para esta solução foram apresentados os seguintes resultados:

i. Conjunto de dados pequeno – Acurácia: 98%
ii. Conjunto de dados grande – Acurácia: 80%
iii. Conjunto de dados muito grande – Acurácia: 87%

Com base nesses resultados a média de acurácia foi de 88,30%.

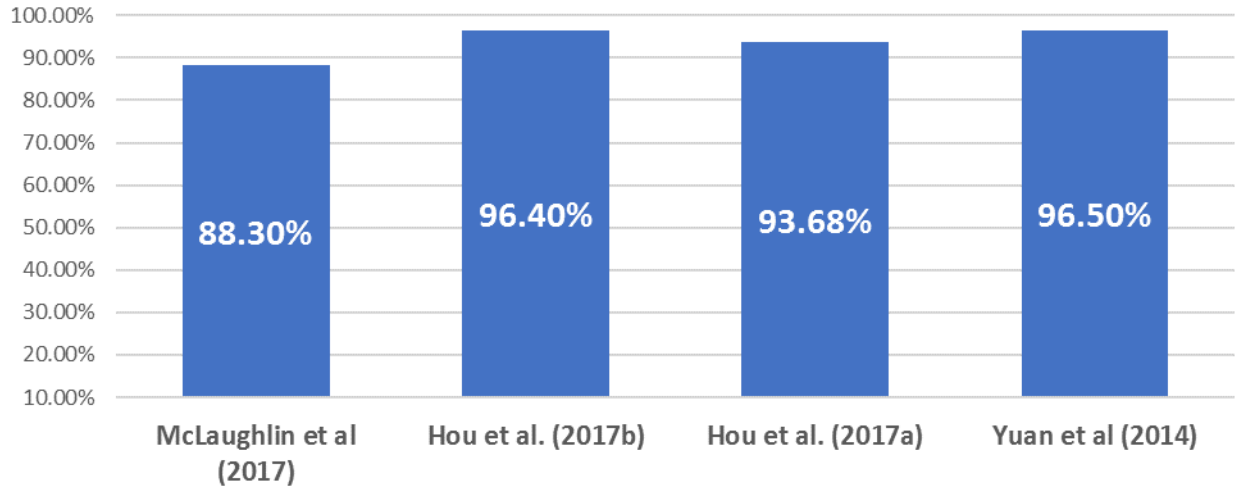
A pesquisa proposta por Hou et al. (2017a), tem como objetivo realizar a detecção de *malwares* conhecidos e desconhecidos (novos) em sistema operacional *android*, utilizando o *Stacked AutoEnconder*, com utilização de 3 mil exemplos de aplicativos (sendo metade deles maliciosos e a outra metade não maliciosos) da *Comodo Cloud Security Center*. O estudo obteve resultado de 93,68% de acurácia.

No trabalho de Hou et al. (2017b) os autores, aprimoraram seus estudos no assunto e escreveram um novo artigo, o qual utiliza *Stacked AutoEnconder* assim como no artigo anterior, porém acrescentando o *Deep Belief Network*. Os pesquisadores aumentaram tamanho da base de dados passando para uma coleção

de 5 mil exemplos reais (2.500 maliociosos e 2.500 não maliociosos) também da *Comodo Cloud Security Center*. Percebe-se que este estudo foi uma evolução do trabalho já realizado anteriormente, pois o percentual de detecções assertivas subiu para 96,40%.

E finalizando este tópico, o estudo realizado por Yuan et al (2014), também possui a finalidade de detectar malwares em dispositivos móveis. Utilizou-se o *Deep Belief Network* como algoritmo, com um dataset contendo aplicativos maliciosos extraídos do *Contagio Mobile* (250 exemplos) e aplicativos não maliciosos extraídos da *Google Play Store* (250 exemplos). E apresentou resultado de 96,50% de acurácia.

Figura 3 – Acurácia para Detecção de Malware (Mobile)



Fonte: Elaborado pelo autor

4.2.2 Resultados para detecção de intrusão

Foram selecionados 4 estudos relacionados a detecção de intrusão em sistemas e redes, os quais terão seus resultados brevemente descritos logo abaixo.

Javaid et al. (2016) apresenta em seu trabalho, uma abordagem baseada em *Deep Learning*, mais especificamente a técnica *Self-taught Learning* para detecção de intrusão. O conjunto de dados utilizados foi o NSL-KDD. Com esta solução os pesquisadores alcançaram um acurácia de 98%.

O segundo trabalho relacionado a detecção de intrusão, realizado por Tian e Li (2017a), propõe um sistema de detecção baseado no uso conjunto de *Deep Belief Network* e o algoritmo de aprendizado de máquina *Random Forest*, e utilizando base

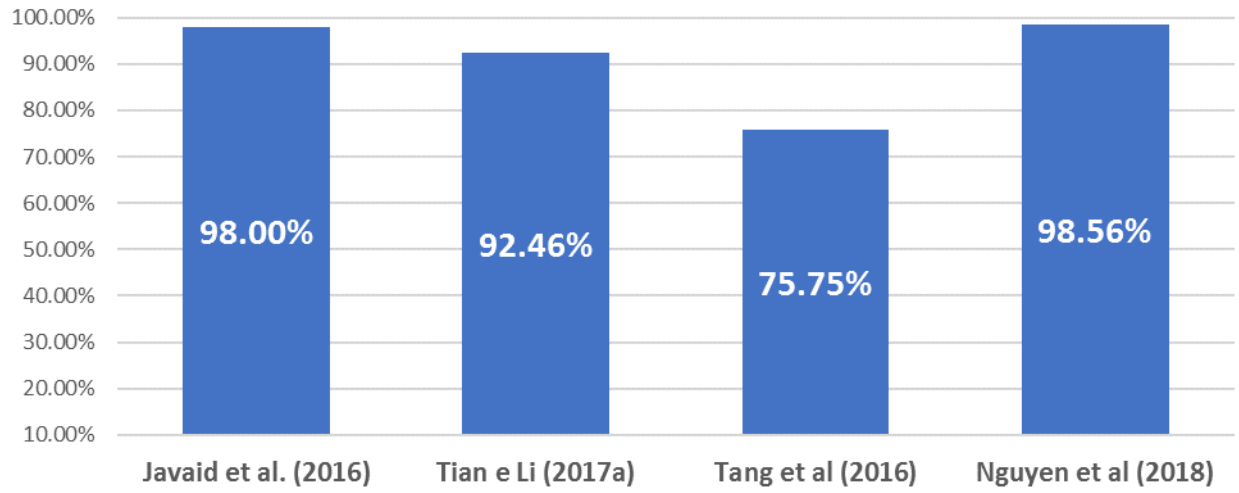
de dados KDD99. Nos experimentos realizados, foram considerados outros métodos, com finalidade de comparação, para que se pudesse analisar qual obteve melhor resultado. Por fim, o modelo descrito no início do parágrafo alcançou um percentual de acerto médio de 92.46%, enquanto os outros chegaram a uma média de:

i. *Deep Belief Network* (sem o uso de *random forest*): 89,82%
ii. *Support Vector Machine* (aprendizado de máquina): 87,06%
iii. *Neural Network* (aprendizado de máquina): 85,09%

O estudo apresentado por Tang et al (2016) foi o que obteve o pior resultado dentre todos os estudos selecionados para esta revisão, utilizando apenas uma simples *Deep Neural Network* e base de dados NSL-KDD, obteve 75,75% de acerto nas detecções realizadas, o que está muito abaixo do praticado em todos outras pesquisas aqui apresentadas.

Por fim, Nguyen et al (2018) apresenta uma solução baseada em *Convolutional Neural Network* para detecção de ataques de negação de serviço. O conjunto de dados utilizado para treino foi o KDD99 e apresentou um percentual muito satisfatório de 98,56% de acerto.

Figura 4 – Acurácia para Detecção de Intrusão



Fonte: Elaborado pelo autor

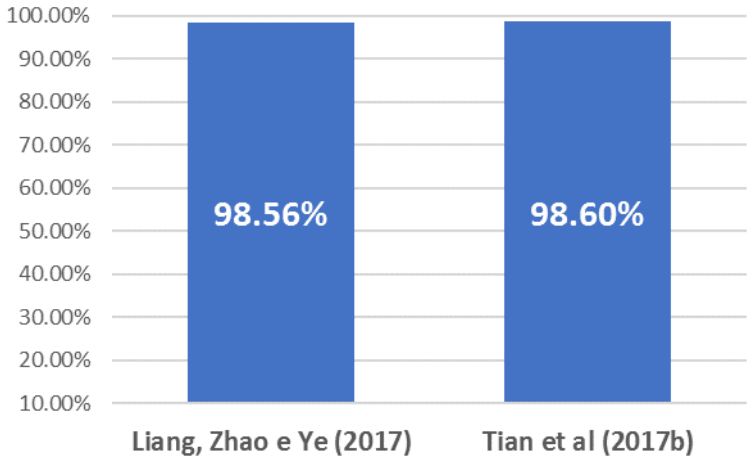
4.2.3 Resultados para detecção de ataques *web*

Este tópico apresenta 2 artigos utilizando abordagem de *Deep learning* para detecção de ataques *web*.

O primeiro artigo deste tópico, demonstra uma solução para detecção de ataques *web*, baseada na detecção de requisições anômalas, utilizando *Recurrent Neural Networks*, que aprende a identificar e diferenciar requisições normais e requisições anormais. A base de dados utilizada para treino da rede neural foi a HTTP CSIC. E a acurácia alcançada foi de 98,56% (LIANG, ZHAO e YE, 2017).

Já no segundo estudo os pesquisadores Tian et al (2017b), utilizaram *Convolutional Neural Network* na solução proposta com o intuito de detectar *web shell* (*scripts* codificados em diferentes linguagens de programação) maliciosos, e a base de dados foi gerada através da coleta pelo *Wireshark* com base em navegação simulada, sendo 3.990 exemplos não maliciosos e 3.691 exemplos maliciosos. O resultado obtido foi de 98,60% de acurácia.

Figura 5 – Acurácia para Detecção de Ataques Web



Fonte: Elaborado pelo autor

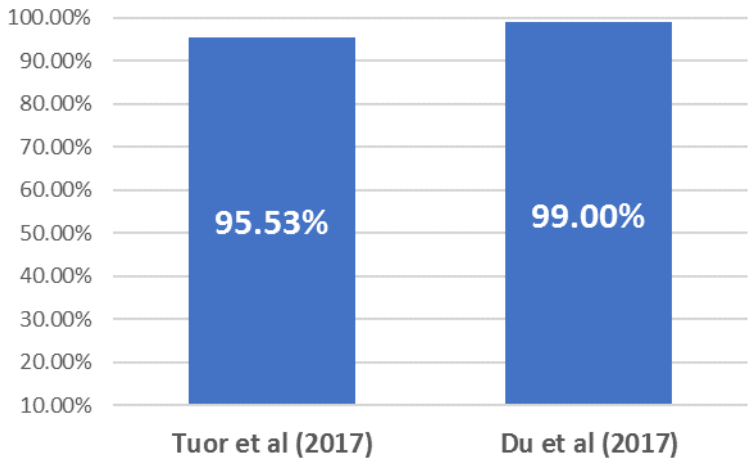
4.2.4 Resultados para detecção de ataques anomalias

Este tópico apresenta 2 artigos utilizando abordagem de *Deep learning* para detecção de anomalias.

O estudo de Tuor et al (2017), tem como objetivo diagnosticar ameaças internas nas redes corporativas, através de detecção de anomalias, para tal, a solução utiliza uma abordagem *online* não supervisionada de *deep learning*, aplicando *Deep Neural Networks* e *Recurrent Neural Networks*. Foi utilizada base de dados CERT *Insider Threat* v6.2 e obteve um percentual de acerto nas detecções de 95,53%.

Outro estudo selecionado para detecção de anomalias foi realizado por Du et al (2017), utiliza *Deep Neural Network* para modelar um *log* do sistema como uma sequência de linguagem natural, o que permite a solução proposta chamada *DeepLog* a aprender automaticamente padrões de *log* de execução normal. A base de dados utilizada para treinar o DeepLog foi HDFS e *OpenStack*. E a acurácia obtida foi de 99%.

Figura 6 – Acurácia para Detecção de Anomalias



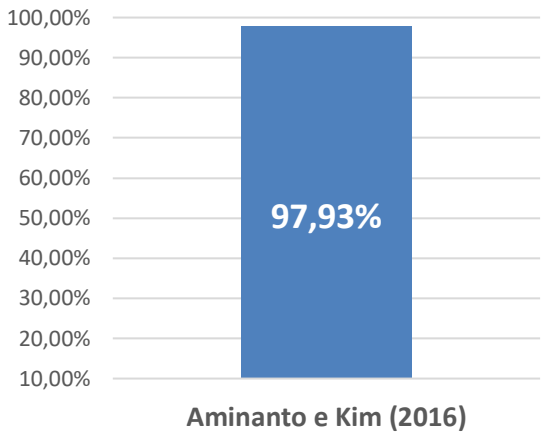
Fonte: Elaborado pelo autor

4.2.5 Resultados para detecção de ataques em redes sem fio

Foi encontrado apenas 1 estudo relacionado a detecção de ataques em redes sem fio, que atendesse todos os critérios de inclusão, portanto não será possível realizar alguma comparação direta com outros estudos.

O estudo desenvolvido por Aminanto e Kim (2016), foca em otimizar a detecção de ataques em redes sem fio por falsificação, para tal foi utilizado *Artificial Neural Network* para seleção de recurso e *AutoEncoder* para classificação baseada no conjunto de dados AWID. O percentual de acurácia alcançado nesta solução foi de 97,93%.

Figura 7 – Acurácia para Detecção de Ataques em Redes sem Fio

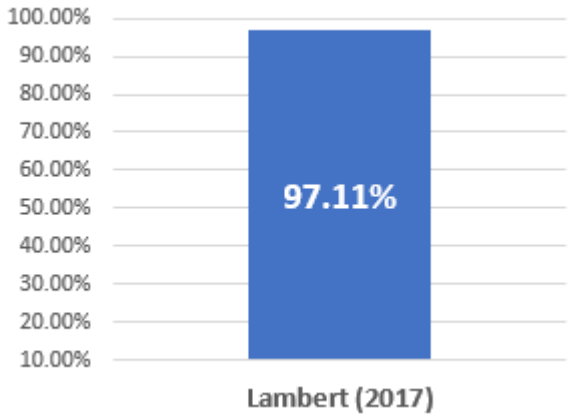


Fonte: Elaborado pelo autor

4.2.6 Resultados para detecção de cyber ataques

O estudo mais abrangente selecionado para esta revisão foi o de Lambert (2017), o qual utiliza um algoritmo de *deep learning* chamado H2O *Deep Learning*, com o intuito de identificar ameaças variadas. A base de dados utilizada foi fornecida pelo Departamento de Tecnologia de Segurança da Informação da Universidade do Norte da Flórida. O percentual médio de acertos da solução ficou em 97,11%, que é muito satisfatório, ainda mais se levarmos em consideração que a solução tem maior "alcance".

Figura 8 – Acurácia para Detecção de Cyber Ataques



Fonte: Elaborado pelo autor

5 CONSIDERAÇÕES FINAIS

Com este trabalho foi possível identificar os avanços na área de segurança da informação, mais especificamente, na área de detecção de cyber ataques aplicando técnicas/algoritmos/métodos de *deep learning*.

Este trabalho buscou reunir informações baseadas na revisão sistemática de literatura, a fim de responder à pergunta: Qual é a efetividade da utilização de algoritmos de *deep learning* para a detecção de *cyber* ataques?

Inicialmente foi necessário entender e descrever os conceitos de uma revisão sistemática, e de IA, AM, DL e de *Cyber* Ataques, de forma a preparar e conduzir este trabalho de maneira adequada e situar o leitor.

Posteriormente foi realizada toda a pesquisa em torno do tema, e encontrar fontes que pudessem então responder à pergunta principal a qual este trabalho se propõe a responder.

Através das análises dos estudos selecionados, pudemos identificar que sim, a aplicação de *deep learning* na detecção de cyber ataques tem se mostrado efetiva, como verificado através da acurácia apresentada nos estudos objeto de avaliação, que em média ficou em torno de 95%.

Portanto é possível concluir que estamos avançando nos trabalhos referente a aplicação de IA para segurança da informação, algo que pode aumentar muito a segurança e tornar mais rápido e fácil a detecção de ataques.

Atualmente existe uma escassez de estudos relacionados ao tema deste trabalho em português, percebe-se a necessidade de realização de pesquisas e experimentos na área para o público falante de português. Além disto, é possível investigar maneiras de aumentar a eficiência e melhorar a performance dos sistemas apresentados.

REFERÊNCIAS BIBLIOGRÁFICAS

ALDERSON, Phil; GREEN, Sally; HIGGINS, Julian. **Cochrane reviewers' handbook 4.2.2**. Chichester, UK: The Cochrane Collaboration, 2004. 234 p.

AMINANTO, Muhamad Erza; KIM, Kwangjo. **Detecting Impersonation Attack in WiFi Networks Using Deep Learning Approach.** WISA 2016: Information Security Applications, Daejeon, p. 136-147, mar. 2017. Disponível em: <https://link.springer.com/chapter/10.1007/978-3-319-56549-1_12>. Acesso em: 11 jun. 2018.

CAKIR, Bugra; DOGDU, Erdogan. **Malware classification using deep learning methods**. ACM, Richmond, mar. 2018. Disponível em: <https://dl.acm.org/citation.cfm?id=3190692>. Acesso em: 11 jun. 2018.

CSO ONLINE. **What is a cyber attack? Recent examples show disturbing trends**. Disponível em: <https://www.csoonline.com/article/3237324/cyber-attacks-espionage/what-is-a-cyber-attack-recent-examples-show- disturbing-trends.html>. Acesso em: 11 jun. 2018.

DAVID, Omid E.; NETANYAHU, Nathan S.. **DeepSign**: Deep learning for automatic malware signature generation and classification. IEEE Xplore, Killarney, out. 2015. Disponível em: <https://ieeexplore.ieee.org/document/7280815/>. Acesso em: 11 jun. 2018.

DENG, Li; YU, Dong. **Deep Learning Methods and Applications**. 7 ed. Redmond, Estados Unidos: Foundations and Trends ® in Signal Processing, 2013. 387 p.

DU, M. et al. **DeepLog: Anomaly Detection and Diagnosis from System Logs through Deep Learning**. ACM, Dallas, p. 1285-1298, out. 2017. Disponível em: <https://dl.acm.org/citation.cfm?id=3134015>. Acesso em: 11 jun. 2018.

GARTNER. **5 trends in cybersecurity for 2017 and 2018**. Disponível em: <https://www.gartner.com/smarterwithgartner/5-trends-in- cybersecurity-for-2017-and-2018/>. Acesso em: 11 jun. 2018.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. 1 ed. Cambridge, MA: MIT Press, 2016. 775 p.

GOOGLE SCHOLAR. **Deep Learning for Cyber Security**. Disponível em: <https://scholar.google.com.br/scholar?hl=pt-BR&as_sdt=0%2C5&as_ylo=2010&as_yhi=2018&q=deep+learning+for+cyber+security&btnG=>. Acesso em: 04 jun. 2018.

GÉRON, Aurélien. **Neural Networks and Deep Learning**. 1 ed. [S.L.]: O'Reilly Media, Inc., 2017.

H2O.AI. **Deep Learning (Neural Networks)**. Disponível em: <http://docs.h2o.ai/h2o/latest-stable/h2o-docs/data- science/deep-learning.html>. Acesso em: 11 jun. 2018.

HAYKIN, Simon. **Neural Networks and Learning Machines**. 3 ed. Ontario, CA: Pearson, 2009. 906 p.

HOU, S. et al. **Deep Neural Networks for Automatic Android Malware Detection**. ACM, Sydney, p. 803-810, jul. 2017a. Disponível em: <https://dl.acm.org/citation.cfm?id=3116211>. Acesso em: 11 jun. 2018.

HOU, S. et al. **Deep4MalDroid:** A Deep Learning Framework for Android Malware Detection Based on Linux Kernel System Call Graphs. IEEE Xplore, Omaha, jan. 2017b. Disponível em: <https://ieeexplore.ieee.org/document/7814490/>. Acesso em: 11 jun. 2018.

II, Glenn M. Lambert. **Security Analytics**: Using Deep Learning to Detect Cyber Attacks. University of North Florida. School of Computing, Florida, mai. 2017. Disponível em: <https://digitalcommons.unf.edu/etd/728/>. Acesso em: 11 jun. 2018.

JAVAID, A. et al. **A Deep Learning Approach for Network Intrusion Detection System**. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), Nova York, p. 21-26, mai. 2016. Disponível em: <https://dl.acm.org/citation.cfm?id=2954780>. Acesso em: 11 jun. 2018.

KITCHENHAM, Barbara; CHARTERS, Stuart. **Guidelines for performing Systematic Literature Reviews in Software Engineering**. EBSE Technical Report, Durham, UK, v. 2, p. 57, jul. 2007. Disponível em: <http://citeseerx.ist.psu.edu/viewdoc/citations?doi=10.1.1.117.471>. Acesso em: 11 jun. 2018.

LIANG, Jingxi; ZHAO, Wen; YE, Wei. **Anomaly-Based Web Attack Detection: A Deep Learning Approach**. ACM, Kunming, p. 80-85, dez. 2017. Disponível em: <https://dl.acm.org/citation.cfm?id=3171594>. Acesso em: 11 jun. 2018.

MCLAUGHLIN, N. et al. **Deep Android Malware Detection**. ACM, Scottsdale, p. 301-3018, mar. 2017. Disponível em: <https://dl.acm.org/citation.cfm?id=3029823>. Acesso em: 11 jun. 2018.

MUNZLINGER, Elizabete; NARCIZO, Fabricio Batista; QUEIROZ, José Eustáquio Rangel De. **Sistematização de revisões bibliográficas em pesquisas da área de IHC**. Sociedade Brasileira de Computação, Cuiaba, BR, p. 51-54, nov. 2012. Disponível em: <https://dl.acm.org/citation.cfm?id=2400099>. Acesso em: 11 jun. 2018.

NGUYEN, S. et al. **Design and implementation of intrusion detection system using convolutional neural network for DoS detection**. ACM, Phu Quoc Island, p. 34-38, fev. 2018. Disponível em: <https://dl.acm.org/citation.cfm?id=3184089>. Acesso em: 11 jun. 2018.

NORTON. **Malware**. Disponível em: <https://us.norton.com/internetsecurity-malware.html>. Acesso em: 11 jun. 2018.

OWASP. **Top 10-2017 Top 10**. Disponível em: <https://www.owasp.org/index.php/Top_10-2017_Top_10>. Acesso em: 11 jun. 2018.

RHODEA, Matilda; BURNAPA, Pete; JONES, Kevin. **Early-Stage Malware Prediction Using Recurrent Neural Networks**. Elsevier, Cardiff, ago. 2017. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0167404818305546>. Acesso em: 11 jun. 2018.

RSA CONFERENCE. **Network Intrusion: Methods of Attack**. Disponível em: <https://www.rsaconference.com/blogs/network-intrusion-methods-of-attack>. Acesso em: 11 jun. 2018.

RUSSELL, Stuart; NORVIG, Peter. **Artificial Intelligence**: A Modern Approach. 3 ed. Nova Jersey, EUA: Pearson, 2012. 1132 p.

SEGARAN, Toby. **Programming Collective Intelligence**. 1 ed. Sebastopol: O'Reilly, 2007. 334 p.

SUTSKEVER, Ilya. **TRAINING RECURRENT NEURAL NETWORKS**. Universidade de Toronto, Toronto, CA, p. 93, jan. 2012. Disponível em: <http://www.cs.utoronto.ca/~ilya/pubs/ilya_sutskever_phd_thesis.pdf>. Acesso em: 11 jun. 2018.

TANG, T. A. et al. **Deep learning approach for Network Intrusion Detection in Software Defined Networking**. IEEE Xplore, Morocco, dez. 2016. Disponível em: <https://ieeexplore.ieee.org/document/7777224/>. Acesso em: 11 jun. 2018.

TIAN, Jingjing; LI, Ping'An. **An Intrusion Detection Algorithm of Dynamic Recursive Deep Belief Networks**. ACM, Nova York, p. 180-183, dez. 2017a. Disponível em: <https://dl.acm.org/citation.cfm?id=3176717&dl=ACM&coll=DL>. Acesso em: 11 jun. 2018.

TIAN, Y. et al. **CNN-Webshell**: Malicious Web Shell Detection with Convolutional Neural Network. ACM, Kunming, p. 75-79, dez. 2017b. Disponível em: <https://dl.acm.org/citation.cfm?id=3171593&dl=ACM&coll=DL>. Acesso em: 11 jun. 2018.

TUOR, A. et al. **Deep Learning for Unsupervised Insider Threat Detection in Structured Cybersecurity Data Streams**. The AAAI-17 Workshop on Artificial Intelligence for Cyber Security WS-17-04, Washington, out. 2017. Disponível em: <https://arxiv.org/abs/1710.00811>. Acesso em: 11 jun. 2018.

UCHÔA, Joaquim Quinteiro. **Segurança computacional**: segurança em servidores linux em camadas. Universidade Federal de Lavras, Lavras, BR, p. 57, jan. 2006. Disponível em: <http://repositorio.ufla.br/handle/1/9307>. Acesso em: 11 jun. 2018.

WANG, W. et al. **Malware traffic classification using convolutional neural network for representation learning**. IEEE Xplore, Da Nang, abr. 2017. Disponível em: <https://ieeexplore.ieee.org/document/7899588/>. Acesso em: 11 jun. 2018.

YUAN, Z et al. **Droid-Sec**: Deep Learning in Android Malware Detection. ACM, Chicago, p. 371-372, ago. 2014. Disponível em: <https://dl.acm.org/citation.cfm?id=2631434>. Acesso em: 11 jun. 2018.