

Deep Learning na identificação de incêndios florestais

João Carlos Sampaio Maldonado Costa Mendes

Engenharia Industrial

Novembro, 2020

Deep Learning na identificação de incêndios florestais

João Carlos Sampaio Maldonado Costa Mendes

Relatório da dissertação apresentada à **Escola Superior de Tecnologia e Gestão do Instituto
Politécnico de Bragança**

Para obtenção do Grau de Mestre em
Engenharia Industrial
Área de especialização em Engenharia Mecânica
sob orientação de:
Professora Doutora Ana Isabel Pereira

Novembro, 2020

A Escola Superior de Tecnologia e Gestão não se responsabiliza pelas opiniões expressas neste relatório.

Agradecimentos

Esta etapa não teria sido concretizada sem a colaboração, direta ou indireta, auxílio e dedicação de várias pessoas ao longo de todo este percurso, sendo esta a oportunidade para agradecer a todas essas pessoas que contribuíram para a minha chegada até aqui não poderia deixar de agradecer:

Em primeiro lugar a pessoa que tornou esta dissertação possível, quer com as melhores ideias de projetos, quer com a sua disponibilidade constante, professora doutora Ana Pereira um muito obrigado não só pela oportunidade de trabalhar consigo novamente como pelo conhecimento passado ao longo dos anos.

Um agradecimento especial aos colaboradores desta dissertação Thadeu Brito e Dr. José Lima pela constante disponibilidade de esclarecimento de dúvidas e partilha de informação e conhecimento. Sem esquecer a todo o corpo docente desta instituição que sempre me acolheu da melhor maneira possível.

À família pelos sacrifícios efetuados, horas perdidas e preocupações para atingir este objetivo, pelo constante apoio e ajuda. Um grande agradecimento a minha namorada pelo permanente incentivo e preocupação.

Um enorme agradecimento a minha outra família, os amigos, que esta cidade me proporcionou e que fizeram desta etapa um percurso memorável, em especial a casa do povo por ser o nosso refúgio de todos os bons e maus momentos, e aos seus fundadores que prevaleceram na minha memória enquanto esta perdurar.

A todos um Muito Obrigado.

Resumo

O problema dos incêndios florestais, é uma constante mundial visto ainda não se ter encontrado uma solução suficientemente eficaz para o seu combate e erradicação. Um dos grandes problemas deste tema no nosso país é o facto da floresta portuguesa possuir uma vegetação muito densa, o que limita o uso de sensores de imagem que segundo alguns autores são a melhor opção para a deteção de ignições florestais.

Uma vez que esta opção de vigilância é excluída e considerando que os primeiros 20 minutos são essenciais para minimizar os danos causados pelo incêndio, deverão ser estudadas e implementadas técnicas mais inovadoras de maneira a combater este flagelo.

Para isso, o projeto SAFe apresenta como solução um conjunto de operações inovadoras que minimizam o tempo de identificação de ignições florestais contribuindo assim, em última instância, para o desenvolvimento da região de Trás-os-Montes. Assim, pretende-se implementar um conjunto de sensores na floresta, onde serão recolhidos os dados referentes as variáveis mais relevantes para a identificação de ignições florestais, uma vez recolhidos esses dados serão analisados através de um sistema de inteligência artificial e por fim serão enviados os alertas as entidades competentes.

O protótipo da caixa de recolha de dados é composto por vários tipos de sensores, nomeadamente sensores de deteção de chama, sensores de temperatura, humidade do solo e do ar e sensores de radiação ultravioleta. O processo de transmissão deste tipo de sensores é assegurado através da comunicação LoRa, uma vez que existe a necessidade de transmitir grandes quantidades de dados (os sensores recolhem dados constantemente em intervalos de dois minutos).

Este sistema permitirá obter um controlo instantâneo, facilitando assim a tarefa de processamento e posterior alerta dentro da janela de tempo necessária para minimizar os danos causados pelo incêndio. O uso da inteligência artificial, mais propriamente do algoritmo *Deep Learning* é uma opção neste caso devido a quantidade de dados obtida no dia a dia, produzindo assim um sistema que se irá tornar mais inteligente a cada hora que esteja a funcionar, devido ao treino constante, o que resultara num *software* mais fiável dia após dia.

Palavras-chave: Inteligência artificial. Vigilância. Ignições Florestais.

Abstract

The problem of forest fires is a constant worldwide since a sufficiently effective solution has not yet been found. One of the major problems of this topic in our country is the fact that the Portuguese forest has a very dense vegetation, which limits the use of image sensors that, according to some authors, are the best option for the detection of forest ignitions.

Since this surveillance option is excluded and considering that the first 20 minutes are essential to minimize the damage caused by the fire, more innovative techniques should be studied and implemented.

For this, the SAFe project presents a solution that carries out a set of innovative operations that minimize the time to identify forest ignitions and contribute to the development of the Trás-os-Montes region. Thus, it is intended to implement a set of sensors in the forest, where the data will be collected regarding the most relevant variables for the identification of forest ignitions, once these data are collected, they will be filtered through an artificial intelligence system and finally the data will be sent. responsible entities.

The prototype of the data collection box is composed of several types of sensors, namely flame detection sensors, temperature sensors, soil and air humidity and ultraviolet radiation sensors. The transmission process of this type of sensors is ensured through LoRa communication, since there is a need to transmit large amounts of data due to the fact that the sensors collect data within two minutes.

This system will allow you to obtain instant control, thus facilitating the task of processing and subsequent alert within the time window necessary to minimize the damage caused by the fire. The use of artificial intelligence, more specifically the Deep Learning algorithm is an option in this case due to the amount of data obtained on a daily basis, thus producing a system that will become more intelligent every hour it is working, due to constant training, which will result in more reliable software day after day.

Keywords: Artificial Intelligence. Surveillance. Forest Ignitions.

Índice

Agradecimentos	iv
Resumo	v
Abstract.....	vi
Índice	vii
Índice de figuras	ix
Índice de tabelas	x
Capítulo 1 - Introdução.....	1
1.1. Motivação	2
1.2. Objetivos.....	2
1.3. Estrutura do relatório	3
Capítulo 2 - Identificação de Ignições Florestais	4
2.1. Caracterização do problema	4
2.2. Tecnologias usadas para a identificação de ignições florestais	5
2.2.1. Sensores óticos e câmaras digitais.....	5
2.2.2. Drones.....	6
2.2.3. Satélites.....	6
2.2.4. Redes de sensores sem fios.....	7
Capítulo 3 - Algoritmos de Inteligência Artificial.....	8
3.2. <i>Big Data</i>	8
3.4. Machine learning	9
3.5. Deep Learning	11
3.5.1. Redes Neurais	11
3.5.2. Tipos de rede	14
Capítulo 4 - Módulos sensorizados SAFe	16
4.1. Solução proposta	16

4.2. Sensores utilizados	18
4.2.1. Sensor de chama	18
4.2.2. Sensor de humidade do solo	19
4.2.3. Sensor de temperatura e humidade do ar	19
4.2.4. Sensor UV	19
4.3. Implementação do algoritmo <i>Deep Learning</i>	20
4.3.1. Identificação dos alertas	22
Capítulo 5 - Resultados obtidos e sua discussão	23
5.1. Simulação I.....	24
5.1.1. Caracterização dos dados.....	25
5.1.2. Resultados obtidos.....	27
5.2. Simulação II.....	37
5.2.1. Caracterização dos dados.....	37
5.2.2. Resultados obtidos.....	39
5.3. Simulação III	42
5.3.1. Caracterização dos dados.....	42
5.3.2. Resultados obtidos.....	44
Capítulo 6 - Conclusões e trabalhos futuros	46
6.1. Conclusões.....	46
6.2. Trabalhos futuros.....	47
Referências Bibliográficas.....	49
Anexos.....	52

Índice de figuras

Figura 3.1- Esquema de uma rede neuronal artificial (Bezerra, 2016).....	13
Figura 3.2- Funções de ativação mais comuns (Bezerra, 2016).....	14
Figura 4.1 - Área de aplicação dos sensores (Brito, et al., 2020).....	17
Figura 4.2- Protótipo da rede implementada (Brito, et al., 2020).....	17
Figura 4.3- Protótipo do módulo de recolha de dados (Brito, et al., 2020).....	20
Figura 5.1- Primeira simulação em laboratório.	24
Figura 5.2- Condições referentes a segunda simulação.....	24
Figura 5.3- Protótipo implementado no campus do IPB (Brito, et al., 2020).....	24
Figura 5.4 - Esquema da primeira simulação.	25
Figura 5.5- Sensores de chama em laboratório.	25
Figura 5.6- Desvio padrão em função da proximidade da chama com os sensores.	27
Figura 5.7- Variação de tamanho de 10% da base de dados de treino.	29
Figura 5.8- Variação de tamanho de 50% da base de dados de treino.	30
Figura 5.9- Variação de constituição número 1.....	32
Figura 5.10- Variação de constituição número 2.....	33
Figura 5.11- Variação de constituição número 4.....	34
Figura 5.12- Alerta final primeira estratégia.	35
Figura 5.13- Alerta final segunda estratégia.....	36
Figura 5.14- Dados referentes a segunda simulação.	38
Figura 5.15- Alerta final da simulação II, 10% da base de dados de treino.....	40
Figura 5.16- Alerta final da simulação II, 50% da base de dados de treino.....	41
Figura 5.17- Dados referentes a terceira simulação.	43

Índice de tabelas

Tabela 1- Caracterização dos dados da primeira simulação	26
Tabela 2- Caracterização dos dados referentes a utilização de 10% de valores de treino	30
Tabela 3- Caracterização dos dados referentes a utilização de 50% de valores de treino	31
Tabela 4- Resumo dos alertas primeira estratégia	36
Tabela 5- Resumo dos alertas segunda estratégia	37
Tabela 6- Caracterização dos dados da segunda simulação	39
Tabela 7- Caracterização 10% de dados de treino segunda simulação	40
Tabela 8- Caracterização 50% de dados de treino segunda simulação	41
Tabela 9- Caracterização dos dados da terceira simulação	43
Tabela 10- Caracterização 10% de dados de treino terceira simulação	44
Tabela 11- Caracterização 50% de dados de treino terceira simulação	45

Capítulo 1 - Introdução

Com o decorrer dos anos, a problemática dos incêndios florestais tem ganho uma nova dimensão quer na opinião pública, quer a nível físico aumentando o número de ignições, o que levou o período de risco de incêndio florestal a ser alargado (Governo, 2017) devido, essencialmente, às alterações climáticas fomentadas pelo aquecimento global. Segundo o trabalho de (Rigo, et al., 2017), a perspetiva de um aumento da temperatura global em cerca de 2°C, acompanhado pela diminuição dos níveis de humidade, irá resultar em níveis de seca mais elevados no meio ambiente, provocando assim um aumento lógico no risco de incêndio florestal. Segundo os mesmos autores, os três países que possuem maior risco são a Espanha, Portugal e a Turquia. Estando em segundo plano países como a Grécia, a região costeira dos Balcãs, regiões central e sul da Itália e ainda do sul de França.

Como é de conhecimento geral, as consequências de um incêndio florestal podem ser devastadoras quando este não for combatido na sua fase inicial, como sugere (Félix & Lourenço, 2017), tal como aconteceu em Portugal no ano de 2017. Neste ano, estima-se que tenham ardido em Portugal cerca de 442 mil hectares, o maior valor desde o ano de 2007 (ICNF, 2017).

Assim sendo, a realização de um sistema automático de deteção de incêndios florestais, preferencialmente na sua fase inicial, seria a solução ideal para este problema. Tendo em conta (Alves, 2018), a utilização de sistemas de videovigilância empregues em Portugal não produziram os efeitos esperados sendo que os sistemas de deteção maioritários continuavam a ser as fontes humanas, nomeadamente as denúncias por parte dos populares e os postos de vigia. De forma a facilitar a deteção de incêndios florestais o presente estudo visa apresentar um novo modo de vigiar as nossas florestas de uma maneira menos dispendiosa e mais eficiente, passando esta pela implementação de várias caixas de recolha de dados para posterior tratamento utilizando as mais recentes vertentes de inteligência artificial.

1.1. Motivação

O primeiro contacto com o projeto SAFE surgiu no âmbito da unidade curricular de inovação baseada em desafios frequentada no passado ano. O projeto SAFE visa criar e implementar um conjunto de operações inovadoras que minimizem a ocorrência de incêndios florestais, monitorizem a fauna e contribuam para o desenvolvimento da região de Trás-os-Montes. Este tema captou bastante a minha atenção, visto que sou residente nesta região, e acima de tudo, devido ao facto de ser um tema bastante complexo e desafiante. Com a conclusão do projeto da unidade curricular “Inovação baseada em desafios”, e após os bons resultados obtidos no mesmo, surgiu a hipótese de elaborar a tese em parceria com o projeto SAFE, dando então continuidade ao projeto iniciado no ano transato.

1.2. Objetivos

Dentro do contexto supracitado, podemos referir como principal objetivo desta dissertação o desenvolvimento do software de tratamento de dados que nos permita reduzir ao mínimo o número de falsos alertas, bem como o tempo de identificação da ignição florestal.

Para este objetivo se tornar exequível, será necessário inicialmente serem atingidos outros objetivos nomeadamente, uma revisão bibliográfica abrangente que nos permita conhecer melhor a problemática e as soluções possíveis, a construção e organização dos dados recolhidos pelos sensores de maneira a criar uma base de dados suficientemente extensa para treinar a nossa rede e por fim aplicar o método em questão comparando o seu resultado com outros similares.

Esta base de dados será composta por várias entradas, detalhadamente dados referentes a quatro tipos de sensores, sensores de chama, sensores de medição de humidade do solo e do ar, sensores de medição ultravioleta e ainda sensores de medição de temperatura.

Uma vez obtidos os dados necessários, será então levada a cabo a sua organização e compreensão, possibilitando assim a aplicação do método em questão para a identificação e alerta de focos de incêndio.

1.3. Estrutura do relatório

De maneira a organizar este documento, o mesmo foi dividido em diferentes capítulos referentes a diferentes temas.

O primeiro capítulo possui um carácter introdutório que pretende enquadrar, o problema proposto, apresentando um pequeno resumo do estado da arte, sendo ainda apresentada a motivação e conseqüentemente os objetivos que se esperam concretizar com esta dissertação.

O segundo capítulo é referente à identificação de ignições florestais, possuindo a caracterização do problema e abordado também as tecnologias em uso para resolver esta problemática.

O terceiro capítulo é alusivo aos algoritmos de inteligência artificial possuindo uma pequena introdução ao uso desta na identificação de ignições florestais, sendo posteriormente é feita uma introdução a todos os fundamentos teóricos necessários para uma melhor compreensão do tema.

O quarto capítulo é dedicado aos módulos sensorizados do SAFe. Este capítulo tem como objetivo descrever as técnicas e os sensores usados para a aquisição dos dados, demonstrando também como foi implementado o algoritmo para a previsão de dados e criação de alertas no software *Matlab*.

No quinto capítulo são caracterizados os dados referentes a cada simulação e posteriormente apresentados os seus resultados, neste capítulo são ainda efetuadas as principais análises e discussões dos resultados em causa.

O sexto e último capítulo, é referente as principais conclusões resultantes do desenvolvimento deste trabalho, bem como sugestões de trabalhos futuros.

Capítulo 2 - Identificação de Ignições Florestais

Neste capítulo será exposta uma caracterização dos incêndios florestais, nomeadamente em Portugal, dando a conhecer, com detalhe, esta temática e explicando sucintamente quais as maiores dificuldades no combate às ignições florestais. Posteriormente, será feita uma introdução às tecnologias em uso para a deteção de ignições florestais.

2.1. Caracterização do problema

O problema dos incêndios florestais, é uma constante mundial visto ainda não se ter encontrado uma solução suficientemente eficaz para o seu combate e irradicação. Tratando-se de um tema em constante discussão e como tal, existem várias técnicas aplicadas no terreno que serão descritas neste capítulo.

Um dos grandes problemas desta temática é o facto da floresta portuguesa possuir uma vegetação bastante densa, limitando assim o uso de sensores de imagem, que segundo alguns autores, nomeadamente (Verma & Singh, 2020) são a melhor opção para a deteção das ignições florestais. Uma vez excluída esta opção de vigilância e tendo em conta que os primeiros 20 minutos são essenciais para minimizar os danos provocados pelo incêndio (Félix & Lourenço, 2017), uma solução proposta será a utilização dos sensores de identificação de chama, humidade, temperatura e radiação ultravioleta.

Obtendo milhares de dados diariamente, outro dos problemas será o tratamento dos mesmos. Tendo como objetivo a identificação de focos de incêndio através dos dados dos sensores supracitados, existe uma necessidade de comparação entre os dados previstos através do nosso software e os dados realmente lidos pelos sensores, para assim, aquando da ocorrência de alguma discrepância de valores, ser lançado um alerta às entidades competentes. Todo este processo necessita ser quase imediato, uma vez que se pretende minimizar o tempo de resposta na identificação e irradicação de ignições florestais.

2.2. Tecnologias usadas para a identificação de ignições florestais

Tendo em conta a problemática em estudo e as inúmeras tentativas de resolução já propostas, existem já várias técnicas aplicadas no mercado, nomeadamente deteção com recurso a câmaras de vídeo integradas, quer em sistemas fixos quer em sistemas móveis. O uso destas resume-se à utilização de sistemas óticos, como se de um olho humano se tratasse com a vantagem de poderem ser utilizadas 24h por dia. Existem ainda outro tipo de mecanismos tais como drones e a utilização de imagens de satélite para identificar as ignições florestais. Nas seguintes secções serão apresentados alguns sistemas já utilizados por todo o mundo.

2.2.1. Sensores óticos e câmaras digitais

Com o desenvolvimento tecnológico corrente, câmaras e sensores combinados com técnicas de processamento de imagem e vídeo permitiram a implementação de sistemas óticos capazes de detetar incêndios florestais (Alkhatib, 2014). Alguns exemplos são:

Câmaras de vídeo

Este tipo de hardware funciona através da captação de imagens ao espectro visível do fumo durante o dia e ao da chama durante a noite (EUFOFINET, 2012). Em concordância com (Yuan, et al., 2017) as câmaras digitais que possuem um sensor de imagem *charge-coupled device*, são câmaras menos confiáveis para aplicações ao ar livre. O grande inimigo destas câmaras é a possibilidade de existirem falsos alertas, tais como os causados por nevoeiro podendo este ser confundido com o fumo durante o dia e durante a noite as cortinas de fumo escondem as chamas, diminuindo assim a fiabilidade das mesmas. Por outro lado, estas câmaras podem fornecer informações valiosas das chamas como o seu comprimento, a altura sobre o terreno e o ângulo da mesma (de Dios, et al., 2007). Uma das grandes vantagens deste tipo de hardware é o seu preço sendo bastante inferior ao valor de uma câmara de infravermelhos (Yuan, et al., 2016).

Câmaras térmicas

São câmaras que captam o fluxo de calor emitido pelo foco de incêndio durante o seu processo de combustão (EUFOFINET, 2012). Estas imagens podem ser capturadas na total ausência de luz (Yuan, et al., 2017). Para além de o fumo ser visto como transparente, as imagens podem ainda ser usadas para fornecer estimativas da frente do incêndio, posição, taxa de propagação e largura da base do mesmo (de Dios, et al., 2007).

Light Detection and Ranging (LIDAR)

Este tipo de sensores, funciona através de um feixe de radiação laser sobre uma nuvem de fumo de modo a detetar a sua posterior reflexão (Traiche, et al., 2008). O uso deste tipo de sensores permite identificar a localização de um foco de incêndio a partir do fumo. Este sistema, funciona tanto de dia como de noite possuindo um alcance de algumas dezenas de quilómetros (Utkin, et al., 2003) sendo a principal desvantagem deste tipo de sensores o seu custo bastante elevado.

2.2.2. Drones

Contrariamente ao que se possa pensar, os drones existem desde cerca de 1961 e são utilizados pelo serviço florestal dos Estados Unidos da América de forma a conseguir obter informações meteorológicas sobre um incêndio em curso (Wilson & Davis, 1988). Posteriormente foram acoplados a câmaras e outros mecanismos que melhoraram em muito o seu uso na deteção de incêndios, podendo referir-se como principais vantagens da sua utilização, a automação do processo, e a capacidade de percorrer grande áreas mesmo com tempo adverso (Yuan, et al., 2015).

2.2.3. Satélites

A deteção de incêndios através de satélites em órbita passa pela análise da temperatura do planeta, identificando as zonas claramente mais quentes que aquelas em seu redor, tem sido aprimorada com o decorrer dos anos, possuindo agora vários aparelhos para este fim, tais como o *Moderate-Resolution Imaging Spectroradiometer* (MODIS), o *Visible Infrared Imaging Radiometer Suite* (VIIRS) e o *Spinning Enhanced Visible and Infrared Imager* (SEVIRI). O MODIS está acoplado ao satélite Terra e Aqua, possuindo uma resolução espacial de aproximadamente um quilómetro (NASA, 2020). Por sua vez, o VIIRS é usado como complemento ao MODIS encontrando-se a bordo do satélite Suomi

NPP, servindo assim para melhorar a resolução espacial conseguindo obter uma resolução de cerca de 375 metros (Changyong, et al., 2014). Deste modo seria então possível detetar e analisar não só os grandes incêndios mas também incêndios de reduzidas dimensões, porém estes dois instrumentos possuem um grande desvantagem, a resolução temporal, uma vez que, em ambos os casos a periodicidade das imagens é insuficiente para os considerar um método de identificação rápida, servindo apenas para recolha de dados e controlo dos incêndios. Analisando agora o caso do SEVIRI que se encontra a bordo do satélite Meteosat-8, é um aparelho que resolve o problema dos anteriores, garantindo imagens a cada 15 minutos porém apenas possui uma resolução espacial de três quilómetros (Calle, et al., 2006)

2.2.4. Redes de sensores sem fios

Outra forma de deteção de incêndios passa pela monitorização dos parâmetros que têm maior influência na sua deflagração, tais como a temperatura, a humidade, a radiação ultravioleta, entre outros. Para tal é necessário elaborar uma rede de sensores estrategicamente posicionados no espaço a controlar, sendo que estes sensores devem ser autónomos do ponto de vista energético e do ponto de vista da transmissão dos dados recolhidos. Estes dados são enviados para uma estação de tratamento onde irão ser analisados e quando algum destes parâmetros atinge valores pré-definidos como indicadores de incêndio, um alarme é acionado (Alkhatib, 2014).

Capítulo 3 - Algoritmos de Inteligência Artificial

Neste capítulo serão abordadas as principais vantagens da utilização da inteligência artificial na temática das ignições florestais, serão ainda abordados os vários conceitos necessários para uma melhor percepção do tema.

Como foi analisado no capítulo anterior, existem neste momento já alguns métodos para a deteção de ignições florestais no mercado (Alves, 2018), porém para todos existe a necessidade de tratamento de dados, uma vez que, quer seja sob a forma de imagem, quer sob a forma de texto, existe uma quantidade massiva de dados recolhidos. Com este enorme número de dados enorme, surge a possibilidade de introduzir uma das tecnologias com maior desenvolvimento, os sistemas de inteligência artificial.

De uma maneira geral, uma definição bem aceite pela comunidade científica apresenta a inteligência artificial como a arte de implementar numa máquina a possibilidade de realizar tarefas inerentes ao ser humano, mas impensáveis de realizar num computador (Rosa, 2011).

Para facilitar a percepção do tema, serão abordados seguidamente os conceitos associados ao *Big Data*, *machine learning* e *deep learning*.

3.2. *Big Data*

Com o decorrer dos tempos, o ser humano evolui para uma sociedade baseada no conhecimento, isto é, para extrair melhor conhecimento é necessária uma maior quantidade de dados quer seja para treinar os algoritmos de inteligência artificial quer seja para própria informação pessoal a nível de notícias diárias, tendo em conta a quantidade de falsas notícias nos dias de hoje.

O conceito de "*Big Data*" refere-se à massa heterogênea de dados digitais produzidos por empresas e indivíduos cujas características requerem ferramentas específicas e cada vez mais sofisticadas de armazenamento e análise de computadores (Riahi, 2018). Não se trata apenas de lidar com volumes cada vez maiores de dados, mas

também a dificuldade de gerir formatos cada vez mais heterogêneos, bem como dados cada vez mais complexos e interconectados. *Big Data* é um conceito muito difícil de definir com precisão, pois a própria noção de “*Big*” em termos de volume de dados varia de uma área para outra. Não é definido por um conjunto de tecnologias, pelo contrário, define uma categoria de técnicas e tecnologias. Este é um campo emergente e onde constantemente tem aparecido diversas referências científicas com novas propostas, à medida que procuramos aprender como implementar esse novo paradigma e aproveitar o valor, a definição está em constante mutação (Riahi, 2018).

3.4. Machine learning

O *machine learning* é uma subárea da inteligência artificial, que ficou conhecida em 1959 por Arthur Samuel como “campo de estudo que dá aos computadores a habilidade de aprender sem serem explicitamente programados”, esta definição mantém-se correta até aos dias de hoje. Os algoritmos de *machine learning* aprendem automaticamente, de um modo geral, através de grandes quantidades de dados, modificando o seu comportamento através da sua própria experiência, consistindo esta mudança comportamental no estabelecimento de regras lógicas que visam melhorar o desempenho de uma determinada tarefa (de Figueiredo, 2020).

Este tipo de algoritmos pode ser subdividido tendo em conta os seus dados iniciais e a maneira como estes estão organizados, como tal neste exemplo específico iremos abordar quatro divisões possíveis:

- *Supervised learning* (aprendizagem supervisionada);
- *Semi-supervised learning* (aprendizagem semi-supervisionada);
- *Unsupervised learning* (aprendizagem não supervisionada);
- *Reinforcement learning* (aprendizagem por reforço);

Estes tipos de divisões são efetuados devido a organização e modo de treino do algoritmo como se explicará seguidamente caso a caso.

Supervised Learning

Neste tipo de funcionamento, é fornecida uma referência do objetivo a ser alcançado, ou seja, é efetuado um treino com o conhecimento do ambiente. Este tipo de treino consiste em conjuntos de entrada com uma saída esperada. O algoritmo de *machine*

learning extrai a representação do conhecimento a partir desses exemplos, sendo que o objetivo principal é desenvolver uma capacidade de gerar saídas corretas para novas entradas, nunca apresentadas ao algoritmo (Pellucci, et al., 2011).

Resumindo, os modelos que utilizam o *supervised learning* aprendem uma função que mapeie as entradas nas saídas, $y=f(x)$ (de Figueiredo, 2020).

Este tipo de *machine learning* pode ser dividido em dois tipos de problemas:

- Problemas de regressão onde y representa um conjunto finito de valores;
- Problemas de classificação onde y representa um conjunto infinito de valores;

Semi-supervised Learning

No *semi-supervised* contrariamente ao que acontece no *supervised learning* nem todos os dados de entrada possuem uma saída esperada. Os dados classificados neste tipo de *learning* são resumidamente menores aos não classificados, sendo que o modelo escolhido será responsável por organizar os dados e posteriormente elaborar as suas classificações. A utilização deste método tem como objetivo utilizar numa fase precoce os dados classificados para o treino inicial, utilizando seguidamente a aprendizagem com recurso a dados não classificados, permitindo assim melhorar consideravelmente a exatidão do método (Gusmão, 2017).

Unsupervised Learning

Este tipo de *learning* é o mais semelhante ao cérebro humano pois não existe qualquer tipo de classificação para os dados de entrada, promovendo assim uma inferência através de uma função que descreva a estrutura das camadas escondidas ou desconhecidas (Gusmão, 2017).

Resumindo, os algoritmos não recebem qualquer tipo de informação sobre o que aprender, tendo que, para esse efeito, descobrir padrões desconhecidos presentes nos dados fornecidos.

Reinforcement Learning

O *reinforcement Learning* possui uma abordagem completamente diferente dos outros tipos, não existindo dados de treino sejam eles rotulados ou não. O algoritmo neste caso tem um funcionamento semelhante ao ensino presente nos dias de hoje, utilizando um sistema de recompensas e penalizações de maneira a obter uma resposta do sistema

aquando do fornecimento dos dados de entrada, avaliando e decidindo qual será o passo seguinte. Neste funcionamento de “tentativa erro”, o sistema deverá aprender as melhores estratégias para obter o maior número de recompensas possíveis ao longo do tempo (de Figueiredo, 2020).

3.5. Deep Learning

Da família dos algoritmos do *machine learning*, o *Deep learning* é inspirado no processo biológico das redes neurais e está a dominar muitas aplicações, provando assim a sua vantagem quando comparado com os algoritmos de *machine learning* convencionais, devido a sua capacidade de produzir resultados mais rápidos e precisos quando a quantidade de dados é aumentada. Estes algoritmos têm como objetivo a representação hierárquica de alto nível dos dados de entrada por meio de camadas de processamento sequencial numa rede neuronal artificial (Kumar, 2018).

A sua investigação deveu-se essencialmente à neurociência mais especificamente ao estudo do córtex visual, pois quando um estímulo visual chega a retina este percorre uma sequência de regiões do cérebro, onde cada uma destas regiões é responsável por identificar características específicas da imagem. Os neurónios das regiões iniciais são responsáveis por detetar formas geométricas simples na imagem enquanto que os neurónios nas regiões finais têm a função de detetar formas gráficas mais complexas, formas estas compostas pelas formas gráficas simples detetadas na região anterior. Desta forma cada região de neurónios (camadas no caso das redes artificiais) combina padrões detetados pela região imediatamente anterior para formar características mais complexas, dando assim origem as famosas redes neurais (Bengio, et al., 2013). Como já explicitado anteriormente os neurónios de uma rede neuronal artificial são organizados em camadas, sendo as mais simples constituídas apenas por uma única camada, composta por um único neurónio. Estes tipos de redes são bastante limitados pois resolvem apenas processos de decisão binários, porém quando se trata de *deep learning* existe um número desconhecido de camadas dependendo do exemplo que se utiliza.

3.5.1. Redes Neurais

A partir do conceito de *machine learning* foi introduzida uma das definições mais importantes para a compreensão do tema, as redes neuronais.

Desde os anos quarenta, que investigadores das mais diversas áreas tentam elaborar um sistema semelhante ao cérebro humano, no que diz respeito ao processamento de informação. Sendo que por intermédio de Warren McCulloch e Walter Pitts com o aparecimento das redes neurais, esta área de investigação sofreu avanços significativos.

Com o decorrer dos anos, foram-lhe conferidas algumas características semelhantes às encontradas nos neurónios dos seres humanos, tais como a existência de pesos e coeficientes de ponderação diferentes associados às ligações entre as várias unidades de processamento, por sua vez este processamento é distribuído e pode ser paralelizado, por fim a aprendizagem é conseguida através do ajuste dos coeficientes de ponderação (Gusmão, 2017).

A transmissão do sinal de um neurónio a outro, no cérebro humano, é um processo químico bastante complexo onde existe a libertação de substâncias específicas pelos neurónios transmissores. O resultado desta operação traduz-se no aumento ou na depressão do potencial eléctrico no corpo da célula recetora, se este potencial alcançar o limite de ativação da célula, uma ação de potência e duração fixa é transmitido para outros neurónios. De uma forma similar, uma rede neuronal artificial possui um sistema de neurónios artificiais sendo que cada um destes neurónios realiza uma computação com base nas demais unidades com as quais está em contacto (Kumar, 2018).

Os neurónios de uma rede neuronal artificial são organizados em camadas, possuindo conexões entre camadas consecutivas, e sendo estas conexões ponderadas por valores reais denominados de pesos como supracitado. Nesta arquitetura, a rede forma um grafo direcionado, no qual os vértices representam neurónios e as arestas representam os pesos das conexões, sendo que a camada que recebe os dados é chamada a camada de entrada enquanto que a que produz o resultado final é chamada de camada de saída, podendo existir uma sequência de L camadas onde ocorre o processamento interno da rede (Fig. 3.1). Sendo que cada elemento dessa camada é chamado de camada oculta (*hidden layer*)

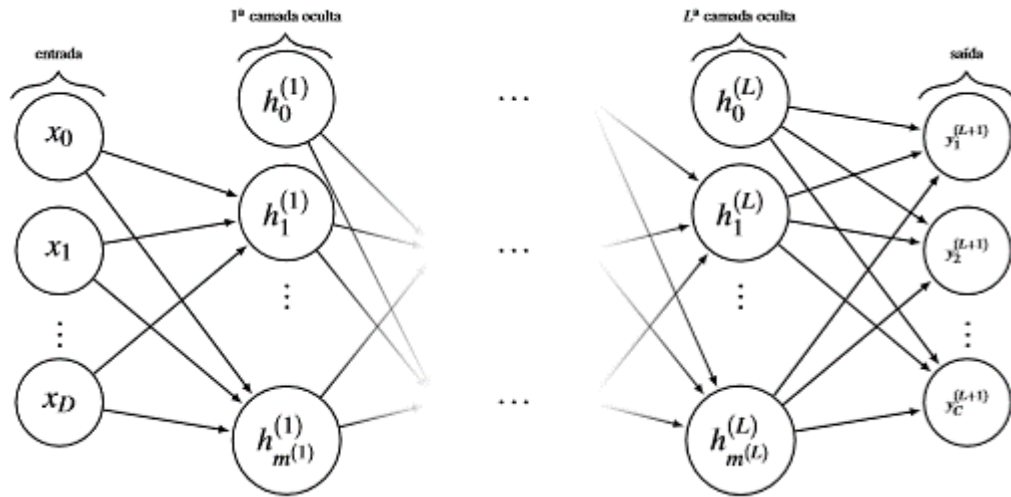


Figura 3.1- Esquema de uma rede neural artificial (Bezerra, 2016).

Cada neurónio de uma camada oculta recebe um vetor $x = (x_0, x_1, \dots, x_n)$ como entrada e calcula uma média ponderada pelos componentes de outro vetor de pesos associados $w = (w_1, w_2, \dots, w_n)$, em que cada componente pode ser interpretado como a força da conexão correspondente. É também comum considerar uma parcela adicional nessa média ponderada, correspondente ao denominado Viés (bias), com peso constante e igual $a+1$ com um único propósito de permitir que a rede neuronal se adapte melhor à função subjacente aos dados de entrada. Sendo que o valor que daí resulta tem o nome de pré-ativação do neurónio (Bezerra, 2016).

$$a(x) = b + \sum_i w_i x_i = b + w^T x \quad (1)$$

Após o cálculo $a(x)$, o neurónio aplica uma transformação não linear sobre esse valor por meio de uma função de ativação. A função composta $h=g \circ a$ produz o que se chama de ativação do neurónio.

$$y = h(x) = g(a(x)) = g\left(b + \sum_i w_i x_i\right) \quad (2)$$

Sendo que existe diferentes alternativas para a função de ativação g , dependendo da tarefa em questão, a função *sigmoide logística* (Equação.3), a *sigmoide hiperbólica tangente* (Equação.4) e a *retificadora linear* (Equação.5) (Bezerra, 2016).

$$g(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

$$g(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (4)$$

$$g(z) = \max(0, z) \quad (5)$$

Como se pode observar na seguinte figura, a exceção da função retificadora linear em $z = 0$, todas são diferenciáveis e estritamente monótonas. Notando ainda que as funções sigmóides possuem assintotas que as limitam quer inferior como superiormente.

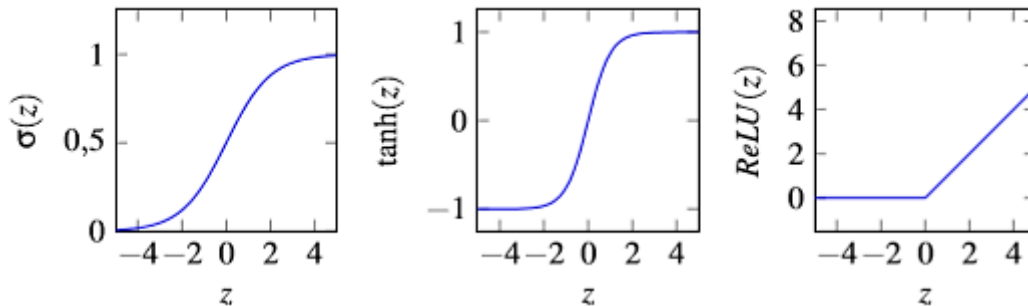


Figura 3.2- Funções de ativação mais comuns (Bezerra, 2016).

As equações supracitadas para a pré-ativação e para a ativação de apenas uma única unidade, podem ser prolongadas para o caso em que existem L camadas ocultas, com vários neurónios em cada camada (Kumar, 2018).

Dentro desta temática, é possível ainda utilizar diferentes classes de redes neurais de acordo com o propósito pretendido, como se irá demonstrar seguidamente.

3.5.2. Tipos de rede

O tipo de redes apresentadas neste subcapítulo são as três classes de redes neurais mais populares no *deep learning*, a saber, redes auto codificadoras, redes convolucionais e redes recorrentes. Estas classes de redes estão na vanguarda das aplicações mais recentes e do sucesso do *deep learning* (Bezerra, 2016).

Redes auto codificadoras

As redes auto codificadoras ou *autoencoders* são uma classe de rede neuronal que pode ser treinada para aprender de forma supervisionada e não supervisionada a partir de um conjunto de dados. Estes tipos de características irão ser úteis quando forem utilizadas em aprendizagem supervisionada tais como, reconhecimento de objetos e outras tarefas relacionadas com a visão computacional. Uma rede auto codificadora codifica os valores de entrada x usando uma função f , seguidamente descodificam os valores codificados $f(x)$ usando uma função g de maneira a criar valores de saída idênticos aos valores de entrada.

O seu objetivo é minimizar o erro de reconstrução entre a entrada e a saída, de

maneira a conseguir aprender os recursos importantes presentes nos dados. Este tipo de redes, atualmente, é maioritariamente utilizado na remoção de ruídos (*denoising*) e redução de dimensionalidade dos dados, podendo assim ser conjugados com as redes neurais convolucionais (CNN) como função preliminar no reconhecimento de imagens (Goodfellow, et al., 2020).

Redes convolucionais

As redes neurais convolucionais (*convolutional neural networks - CNN*) inspiram-se no funcionamento do córtex visual, são normalmente conectadas completamente, ou seja, cada unidade da camada L está conectada a todas unidades da camada $L-1$.

Desde a sua origem que são utilizadas em tarefas de reconhecimento visual, como reconhecimento de manuscritos e deteção de faces. Ganharam popularidade em 2012, quando conseguiram reduzir substancialmente a taxa de erro na competição de reconhecimento de objetos, a *ImageNet* (Bezerra, 2016).

Redes recorrentes

As redes neuronais recorrentes (*Recurrent Neural Networks - RNN*) constituem uma ampla classe de redes cujo objetivo é reconhecer padrões em sequências de dados, como por exemplo texto ou bolsas de valores. Estes algoritmos apresentam uma dimensão temporal, isto é, consideram o tempo e a sequência.

Estes tipos de redes tomam como valores de entrada não apenas o exemplo de entrada atual que recebem, mas também o que receberam anteriormente, possuindo assim duas fontes de entrada, o presente e o passado recente. Essa propriedade proporciona a possibilidade de realizar os cálculos dependente do contexto e aprender dependências de longo prazo: um sinal que é fornecido a uma rede recorrente num instante de tempo t pode alterar o comportamento dessa rede num momento $t+k$, $k>0$ (Goodfellow, et al., 2020).

Capítulo 4 - Módulos sensorizados

SAFe

Este capítulo tem como objetivo descrever as técnicas e os sensores usados para a aquisição dos dados de maneira a identificar rapidamente as ignições florestais, pretende-se ainda demonstrar como foi implementado o algoritmo para a previsão de dados e criação de alertas no software *Matlab*.

4.1. Solução proposta

A proposta de resolução para este tema já se encontra em desenvolvimento há cerca de dois anos, uma vez que se trata de um tema proposto no âmbito do projeto SAFe.

Após a revisão da literatura, concluiu-se que o melhor método de deteção de ignições florestais passa pela implementação de sensores, quer pela questão económica quer pela facilidade de manutenção do sistema. A utilização dos sensores permite não só a deteção dos incêndios florestais, como também no período menos crítico, controlar o crescimento da biomassa presente em determinada localização, possibilitando ainda a monitorização em tempo real das condições atmosféricas da região em estudo. Para tal vão ser utilizados quatro tipos de sensores que serão apresentados seguidamente bem como todas as suas características técnicas.

A aplicação dos módulos de recolha de dados será implementada na região de Bragança, mais propriamente na área relativa à Serra da Nogueira (Fig. 4.1), existindo a necessidade de desenvolver estratégias para a colocação dos sensores, tendo em conta a densidade da floresta, a elevação do terreno e ainda equacionando as áreas que já sofreram ou estão mais propícias a sofrer ignições florestais.

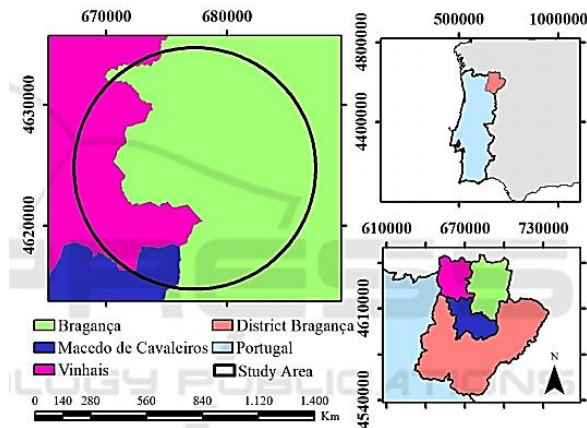


Figura 4.1 - Área de aplicação dos sensores (Brito, et al., 2020).

Tendo em conta que os sensores irão ser aplicados a vários quilómetros de distância do centro de controlo e a necessidade, de transmitir dados constantemente, torna-se fulcral a existência de um sistema de comunicação eficiente. Para tal foi adotada a tecnologia LoRa (Long Range) de modo a garantir a conexão entre os sensores e a central. O sistema de comunicação LoRa trata-se de uma tecnologia de rede de banda larga e de baixo consumo permitindo transmissões de longo alcance, cerca de 10 quilómetros em meio rural, permitindo assim a comunicação dos dados dos sensores (Brito, et al., 2020).

Em síntese iremos ter uma rede conforme se apresenta na seguinte imagem:

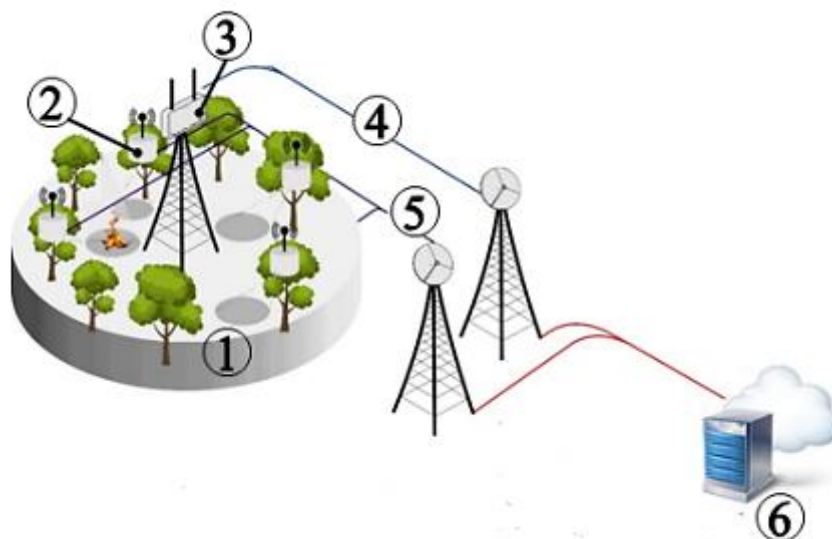


Figura 4.2- Protótipo da rede implementada (Brito, et al., 2020).

Onde:

1. Área em estudo;
2. Modulo de recolha de dados;
3. LoRaWAN Gateway, o dispositivo que recebe os dados de todos os sensores implementados na área em estudo;
4. Comunicação através da internet, que nos facilita a transmissão em direto de dados;
5. Transmissão LoRa;
6. Servidor onde será aplicado o algoritmo de inteligência artificial que nos vai possibilitar a criação dos alertas;

Quando os dados chegam ao servidor, é então implementado o método *deep learning* que nos permitirá, através dos dados recebidos prever os dados esperados. Quando existirem dados esperados será então possível criar os alertas através de comparações entre os dados expectados e os dados realmente obtidos através dos sensores. A implementação deste algoritmo será apresentada mais em detalhe seguidamente, tornando o processo mais explícito.

4.2. Sensores utilizados

Para a realização do projeto foram seleccionados, após um estudo bibliográfico extenso, quatro tipo de sensores. Tratando-se do projeto piloto foram escolhidos sensores para conexão ao Arduíno uma vez que é um hardware experimental permitindo assim uma mais fácil montagem e programação do sistema de recolha de dados.

Os quatro tipos de sensores escolhidos foram sensores de chama, sensores de humidade do solo, sensores de temperatura e humidade do ar e ainda um sensor de medição ultravioleta como será explicado seguidamente (Pereira, et al., 2019).

4.2.1. Sensor de chama

Este sensor, tal como o nome indica é utilizado para detetar qualquer tipo de ignição, sendo sensível ao espectro da chama e ajustável consoante a intensidade a ser registada.

Especificações:

- Espectro de funcionamento: 760nm ~ 1100nm

- Ângulo de funcionamento: 0 - 60 graus
- Tensão de funcionamento: 3.3V ~ 5.3V
- Temperatura: -25 °C ~ 85 °C
- Dimensões: 27,3 mm * 15,4 mm
- Diâmetro da lente: 2.0 mm

4.2.2. Sensor de humidade do solo

O sensor é utilizado para controlar a humidade do solo, sendo constituído por duas partes. Uma parte é utilizada para recolher os dados, ficando posicionada no solo, sendo que o outro consiste num módulo que contém um chip LM393 o qual envia os dados para um microcontrolador.

4.2.3. Sensor de temperatura e humidade do ar

Este sensor contém um módulo digital DHT11 e vai ser utilizado para recolher dados relativos à temperatura e humidade instantânea, devendo este sensor estar em contacto com o ar exterior.

Especificações:

Temperatura

- Resolução: 1 ° C
- Precisão: ± 2 ° C
- Intervalo de funcionamento: 0 ° C ~ 50 ° C

Humidade

- Resolução: 1% RH
- Precisão: ± 5 % RH (0 ~ 50 ° C)
- Intervalo de funcionamento: 20% RH ~ 90% RH (25 ° C)
- Tensão de funcionamento: 3.3 V ~ 5,5 V

4.2.4. Sensor UV

O sensor de medição ultravioleta vai mensurar a radiação sentida, resultando numa informação entre 0-1 V, correspondendo a uma escala de 0-10 UV.

No final da montagem de todos os sensores o protótipo deverá ser semelhante ao apresentado na seguinte imagem:

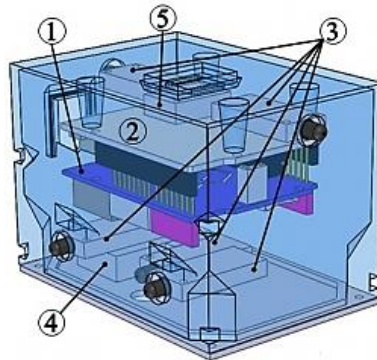


Figura 4.3- Protótipo do módulo de recolha de dados (Brito, et al., 2020).

Sendo os seus componentes representados por:

1. Arduíno Uno;
2. Modulo da comunicação LoRa;
3. Sensores de chama;
4. Sensor de humidade e temperatura do ar;
5. Sensor ultravioleta;

4.3. Implementação do algoritmo *Deep Learning*.

De forma a ser possível implementar o algoritmo foi necessário proceder à investigação dos modelos disponíveis no software *Matlab* (Mathworks, 2020). Após o estudo do mesmo, conclui-se que a melhor solução neste caso, seria a construção de uma *long short-term memory* (LSTM), simplificando, um processo que utiliza a arquitetura da rede neuronal recorrente (RNN) e que nos permite trabalhar com conjuntos de dados.

O funcionamento do código tem por base a importação de uma base de dados, de maneira a ser possível elaborar a previsão tendo em conta os dados originais. Neste caso inicial como os dados se encontravam em formato de texto, foi necessário transformá-los para formato de livro de Excel (.xlsx), de maneira a serem mais facilmente lidos pelo *Matlab*.

Uma vez importados, torna-se necessário selecionar a quantidade de dados que se deseja utilizar para o treino e para o teste do algoritmo, podendo variar este valor de acordo com as nossas necessidades, desde utilizar somente 99% de dados para treino e utilizar os restantes 1% para teste, ou qualquer outro tipo de variação que se deseje. Tendo em conta a separação de dados efetuada, torna-se necessário garantir a utilização exclusiva de valores inteiros utilizando para tal a função *floor* presente no *Matlab*. Seguidamente tornou-se necessário proceder a uma pequena adaptação, através da função transposta, uma vez que por defeito a programação lê linhas e os dados obtidos através dos sensores possuem formato de coluna, criando assim as duas primeiras variáveis o “*datatrain*” e o “*datatest*”.

Com os dados devidamente implementados no algoritmo e munidos das duas primeiras variáveis, torna-se necessário (para evitar uma divergência no momento do treino), padronizar os dados de treino de maneira a obter uma média zero e cuja variação corresponda à unidade em causa. Relembrando que no momento da previsão é igualmente necessário padronizar os dados de teste de maneira a serem iguais aos dados de treino, sendo necessário para tal calcular a média e o desvio padrão dos dados inicialmente.

Tendo os dados já devidamente preparados, torna-se necessário adaptar agora a parte do treino e da arquitetura da rede, neste processo é possível ajustar o número de unidades ocultas, o “*solver*”, o número de épocas em que se deve treinar, os limites de gradiente, a taxa de aprendizagem, entre outras opções. Após encontrar as melhores opções para o treino da rede é necessário que esta treine, o que vai ser conseguido através da função *trainNetwork*, utilizando para tal, as variáveis “*Xtrain*” o “*YTrain*” e todas as variáveis ajustadas anteriormente.

Uma vez finalizado o treino da rede, torna-se necessário codificar a parte da previsão usando para tal a função *predictAndUpdateStat*, prevendo então os valores referentes às variáveis “*Xtrain*” e “*YTrain*” atualizando sempre o estado da rede. Concluída a atualização com os valores previstos, é necessário despadronizar os dados novamente de modo a ser possível apresentar os dados previstos. Para existir uma noção da fiabilidade é necessário proceder ao cálculo do erro quadrático através dos dados não padronizados.

4.3.1. Identificação dos alertas

Possuindo os dados previstos, passar-se-á à identificação dos alertas. Será considerado um alerta individual qualquer valor que se encontre díspar o suficiente dos dados previstos, para tal serão apresentadas seguidamente duas condições distintas que serão devidamente testadas de maneira a perceber qual a mais adequada para o efeito pretendido. Por sua vez o alerta final será composto pelo conjunto de alertas individuais similares aos quatro tipo de sensores de maneira a reduzir os possíveis falsos alertas, melhorando assim a fiabilidade do sistema.

Neste trabalho foram testadas várias condições de alerta, apresentando seguidamente as duas que surtiram melhores resultados:

- A primeira abordagem foi feita através da diferença entre os valores previstos e os valores reais obtidos pelos sensores (Equação.6). Uma vez que se trabalha com a diferença de valores, neste caso, queremos encontrar todos os valores que sejam superiores à média mais o desvio padrão desta diferença, resultando na seguinte condição:

$$x = |\text{Dados. Previstos} - \text{Dados. Reais}|$$
$$\text{Alerta se: } (x > (\text{Média}(x) + \text{Desvio padrão}(x))) \quad (6)$$

- A segunda estratégia compara os valores previstos diretamente com os valores reais (Equação.7) logo, neste caso, o objetivo é encontrar os valores que sejam menores que a média menos o desvio padrão dos dados reais resultando:

$$y = (\text{Dados. Previstos})$$
$$z = (\text{Dados. Reais})$$
$$\text{Alerta se: } (y < (\text{Média}(z) - \text{Desvio padrão}(z))) \quad (7)$$

Salientando que as condições foram aplicadas inicialmente para cada sensor individualmente e ponto a ponto, somente quando estes alertas individuais foram identificados é que foi possível apresentar o alerta final (conjunto de todos os alertas individuais).

Capítulo 5 - Resultados obtidos e sua discussão

No presente capítulo serão expostos os resultados obtidos através do software *Matlab*. Este software foi acionado num computador com um processador *intel core I5-5200 CPU 2.20GHz* com 8Gb de memória RAM. A versão do software utilizada foi a R2019a (9.0.0.341360).

Considerando a situação de atual pandemia que vivemos, a forma de recolha de dados teve de ser alterada. Inicialmente os dados provenientes dos quatro tipos de sensores seriam recolhidos no recinto do IPB quando se procedesse às queimadas sazonais. Uma vez que foi impossível este tipo de recolha, foram elaboradas três simulações, organizadas da seguinte forma:

- Simulação I - simulação elaborada com recurso a materiais simples, num contexto de laboratório recorrendo a uma chama de pequenas dimensões (Fig. 5.1);
- Simulação II - simulação efetuada no exterior de maneira a simular as condições presentes na maior parte do território nacional com presença de biomassa que nos permitisse atingir chamas de pelo menos 40 cm (Fig. 5.2);
- Simulação III - a última simulação é referente aos dados obtidos através das caixas de recolha presentes no recinto do IPB (Fig. 5.3);

A recolha de dados teve de ser adaptada uma vez que não foi possível obter dados dos quatro tipos de sensores, obtendo apenas os dados dos cinco sensores de chama, temperatura e humidade do ar, sendo que na simulação I e II apenas foram utilizados os sensores de chama.

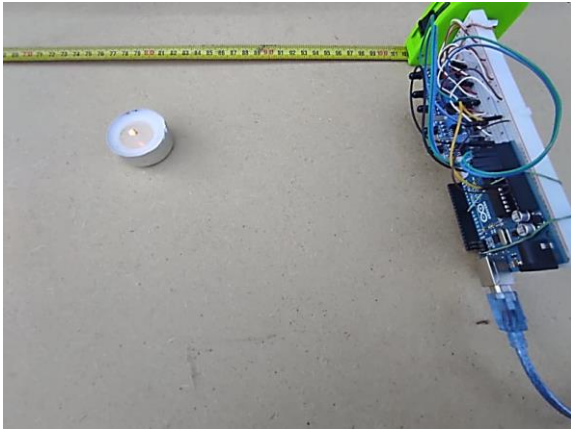


Figura 5.2- Primeira simulação em laboratório.



Figura 5.1- Condições referentes a segunda simulação.



Figura 5.3- Protótipo implementado no campus do IPB (Brito, et al., 2020).

5.1. Simulação I

Na primeira simulação recorreu-se a 5 sensores de chama, com o objetivo de analisar e estudar o comportamento destes. Os dados obtidos são referentes a uma chama de pequenas dimensões (1 a 2 cm), colocada a cerca de um metro dos sensores de chama (Fig. 5.4) que foi sendo aproximada e afastada dos sensores progressivamente, de forma a podermos testar a sua fiabilidade. Podendo a simulação ser dividida em cinco momentos distintos, tendo em conta a presença e a posição da chama, da seguinte maneira:

- Do instante 0 ao instante T_1 - Não existe qualquer tipo de ignição;

- Do instante T_1 ao instante T_2 - A chama é acesa e é aproximada de 10 em 10 cm, em intervalos constantes de tempo, dos sensores;
- Do instante T_2 ao instante T_3 - A chama atinge o ponto máximo de proximidade com os sensores e é afastada novamente de 10 em 10 cm, com os mesmos intervalos de tempo, dos sensores;
- Do instante T_3 ao instante T_4 - A chama atinge o seu ponto original, a cerca de um metro do sensor, mantendo-se nesta posição;
- Instante T_5 - A chama é extinta;

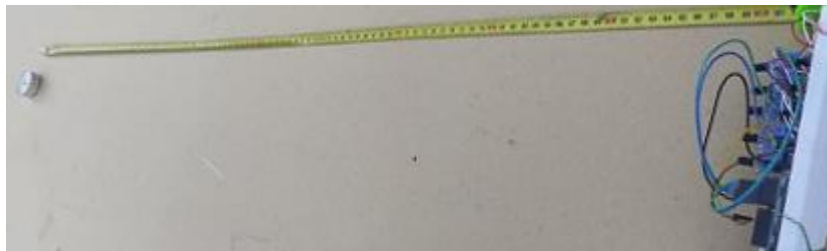


Figura 5.4 - Esquema da primeira simulação.

5.1.1. Caracterização dos dados

Os dados a analisar são obtidos através dos cinco sensores de chama disponíveis, resultando na seguinte figura:

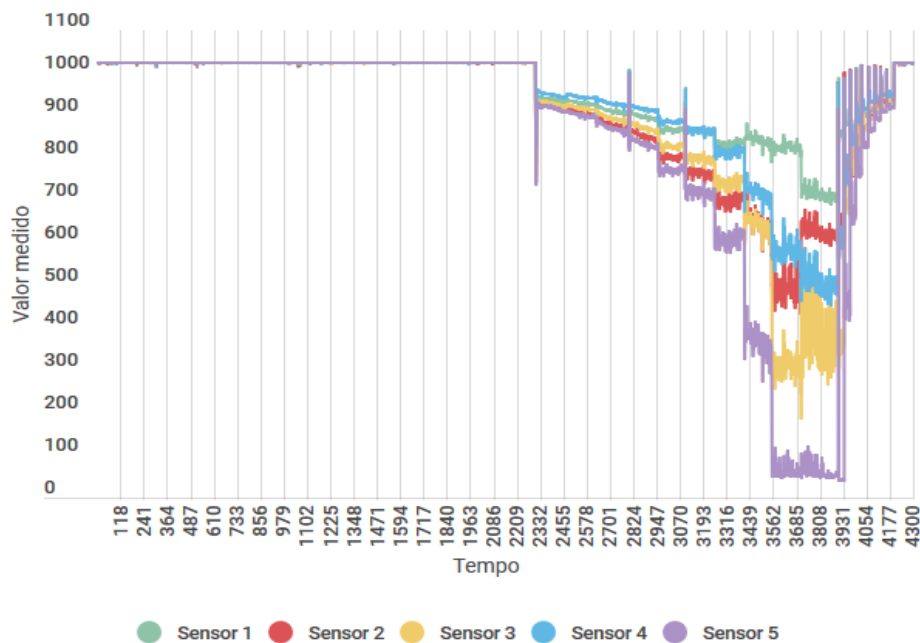


Figura 5.5- Sensores de chama em laboratório.

A Fig. 5.5 apresenta os resultados obtidos através dos cinco sensores em função da unidade de tempo, sendo que, cada unidade de tempo corresponde a 2 segundos. O teste consiste no seguinte protocolo:

- Entre o instante 0 até ao instante 2307 não existe qualquer tipo de ignição;
- No instante 2308 uma chama foi acesa a cerca de 100 cm dos sensores, e posteriormente aproximada 10 cm;
- No instante 2470 a chama encontrava-se a 90 cm dos sensores, prosseguindo o mesmo processo até atingir a distância mínima do sensor cerca de 10 cm, no instante 3710;
- Do instante 3710 ao instante 3860 a chama manteve uma posição constante a 10 cm dos sensores
- A partir do instante 3861 a chama foi afastada, repetindo o mesmo processo, porém aguardando apenas um minuto entre cada 10 cm.
- Do instante 4160 ao instante 4199 a chama manteve-se acesa a cerca de 100 cm dos sensores novamente;
- No instante 4200 a chama foi extinta permanecendo assim até ao instante 4300 quando a simulação terminou;

Para a elaboração do treino da rede foram utilizados os dados dos cinco sensores de maneira a obter um resultado mais fiável, variando as suas médias da seguinte forma:

Tabela 1- Caracterização dos dados da primeira simulação

Sensor nº	1	2	3	4	5	
Sem chama	1022.62	1022.50	1022.58	1022.61	1022.61	
Média a distância de:	100 Cm	961.83	952.55	955.01	970.06	947.71
	90 Cm	956.89	945.55	948.30	966.51	947.71
	80 Cm	912.50	884.20	894.65	929.87	871.70
	70 Cm	899.36	858.78	873.54	916.54	837.52
	60 Cm	872.55	813.15	834.24	889.58	779.52
	50 Cm	864.20	780.94	808.58	870.98	735.25
	40 Cm	842.50	712.8	735.28	805.22	585.21
	30 Cm	845.13	648.16	599.41	700.24	319
	20 Cm	816.07	550.8	362.08	587.01	91.62
10 Cm	716.69	628.3	398.23	512.18	63.49	
Média total	953.78	912.66	898.54	931.63	846.51	
Desvio Padrão	88.77	152.44	197.3	143.5	290.64	

Como se pode verificar na Tabela 1, os cinco sensores apresentam valores similares relativamente aos dados sem chama possuindo um desvio padrão de apenas 0.04, o que permite afirmar que os cinco sensores estão a efetuar leituras próximas, sem qualquer tipo de anomalia.

Uma vez iniciada a ignição da chama, existe um aumento do desvio padrão, mostrando assim a diferença de leituras entre os sensores, este aumento pode dever-se a vários fatores nomeadamente a posição dos sensores em relação à chama, as variações da chama devido à influência do ar, entre outros. Seguidamente é apresentada a Fig. 5.6 que demonstra o comportamento do desvio padrão em função da proximidade da chama dos sensores.



Figura 5.6- Desvio padrão em função da proximidade da chama com os sensores.

Observando a Fig. 5.6 é de fácil percepção que existe um aumento proporcional do desvio padrão com a aproximação da chama até a distância de 20 cm dos sensores, baixando novamente quando a chama se encontra a 10 cm dos sensores. Esta quebra de padrão deve-se ao facto da curta distância a que a chama se encontra dos sensores, resultando assim numa normalização das leituras.

5.1.2. Resultados obtidos

Como foi referido anteriormente existem várias variáveis que podem ser trocadas, modificando assim a maneira de como o algoritmo procede ao treino da rede neuronal,

para se perceber como funcionam este tipo de variáveis é necessário ter em conta quatro conceitos essenciais:

- Base de dados inicial - base de dados constituída na sua totalidade pelos dados adquiridos através dos sensores;
- Base de dados de treino - esta base de dados é constituída por uma percentagem da base de dados inicial, que servirá para identificar o modelo;
- Dados previstos - conjunto de dados previstos obtidos através do algoritmo;
- Base de dados de teste - base de dados reais, dados adquiridos pelos sensores, que têm como objetivo complementar a base de dados de treino de modo a ser possível verificar a veracidade dos dados previstos;

Tendo em conta os conceitos apresentados, a primeira alteração que foi efetuada foi a variação do tamanho da nossa base de dados de treino, variando esta entre dois valores, 10% e 50%.

Esta variação irá influenciar também o tamanho da base de dados de teste e os nossos dados previstos, uma vez que, os dados reais disponibilizados não necessitam de ser previstos, funcionando assim como complementares entre si, isto é, se o tamanho da nossa base de treino for 50% da base de dados inicial somente serão previstos 50% dos dados e por sua vez só irão ser utilizados 50% da base de dados de teste para confirmar a veracidade da previsão.

Uma vez definido o tamanho da nossa base de dados de treino, é ainda possível definir com que valores é que esta é constituída. Podendo a base de dados ser somente constituída por valores sem presença de chama, ou apenas com valores obtidos na presença de uma ignição. Podemos também alterar a constituição da nossa base de dados de teste.

Como será apresentado seguidamente foram testadas quatro diferentes constituições quer para a base de dados de treino quer para a base de dados de teste, variando estas constituições da seguinte forma:

- 1) Base de dados de treino com valores somente sem chama e base de dados de teste com valores somente na presença de chama;
- 2) Base de dados de treino com valores somente na presença de chama e base de dados de teste com valores somente sem chama;
- 3) Base de dados de treino e base de dados de teste constituídas pela totalidade dos valores não existindo qualquer diferenciação de dados com ou sem chama;

4) Treino com a mesma quantidade de valores de chama e sem chama¹;

Por fim e como foi apresentado anteriormente será ainda variada a forma como os alertas são identificados, de modo a surtir os melhores resultados.

Resultados simulação I - Variação do tamanho da base de dados de treino

De modo a que se possa perceber melhor a influência do tamanho da nossa base de dados de treino, irão ser apresentadas duas formas de comparação. Inicialmente serão apresentados os gráficos referentes aos valores previstos e posteriormente uma comparação entre as suas médias e desvios padrões.

Estas comparações serão efetuadas sem qualquer variação da constituição da base de treino, isto é, quer a base de dados de treino quer a base de dados de teste possuem a mesma constituição (variação 3), nomeadamente a totalidade de dados sem qualquer distinção de valores com e sem chama.

Seguidamente serão apresentados os gráficos referentes aos valores previstos, em função da unidade de tempo, mantendo-se esta constante em intervalos de 2 segundos durante toda a simulação. A primeira figura corresponde à variação de tamanho de 10% da base de dados de treino, ou seja, 90% de dados previstos, e a segunda figura é referente aos 50%.

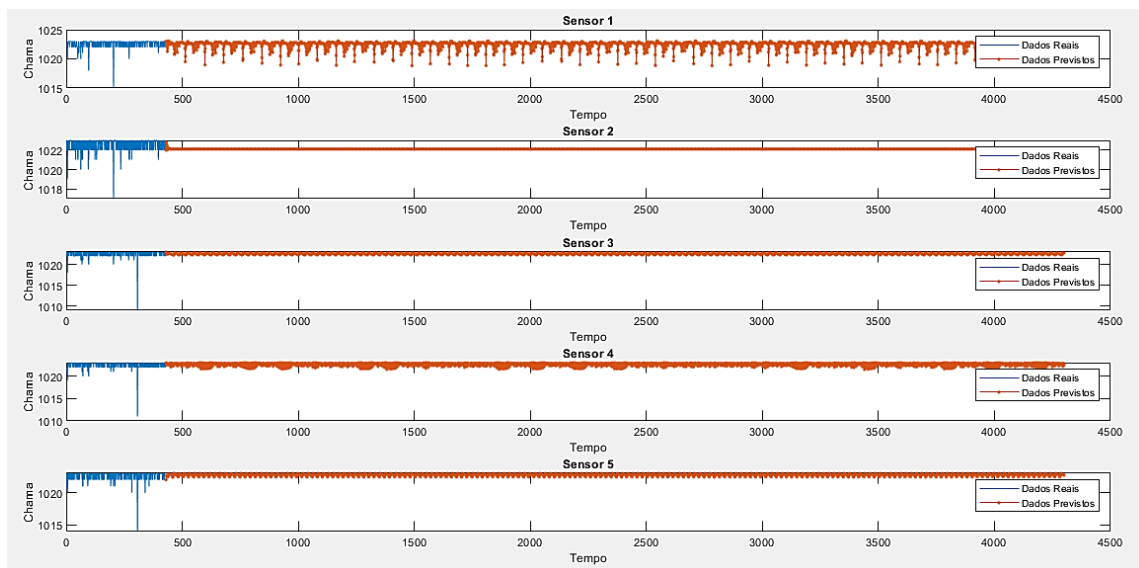


Figura 5.7- Variação de tamanho de 10% da base de dados de treino.

¹ Foram escolhidos 50% de dados na presença de chama e 50% de dados sem chama, precisamente, sendo estes dados a constituição quer da nossa base de dados de treino quer na nossa base de dados real.

Observando a Fig. 5.7 é de fácil percepção que existem diferenças significativas entre as previsões de cada sensor, estando estas diferenças relacionadas com os valores realmente lidos pelos sensores, confirmando assim a variação do desvio padrão entre os sensores. De modo a tornar a comparação mais fácil será apresentado seguidamente o gráfico referente a variação de tamanho de 50% da base de dados de treino:

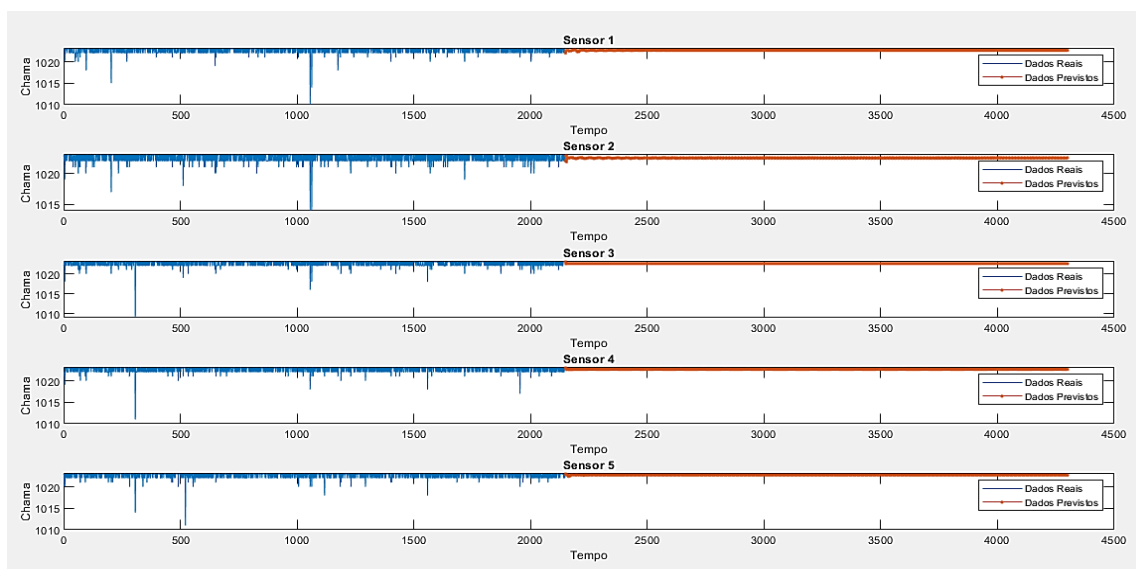


Figura 5.8- Variação de tamanho de 50% da base de dados de treino.

Confrontando as duas imagens relativas às previsões podemos observar que quando se trabalha com 50% da base dados de treino os valores da previsão tendem a fixar num valor único enquanto que, quando se trabalha com 10%, a previsão tende a oscilar entre dois ou mais valores. Seguidamente serão apresentadas as tabelas referentes às figuras anteriores:

Tabela 2- Caracterização dos dados referentes a utilização de 10% de valores de treino

Sensor	Média base de dados treino	Média valores previstos	Desvio Padrão base de dados treino	Desvio Padrão valores previstos
1	1022.6	1022.5	0.7	0.7
2	1022.5	1022.2	0.67	0.28
3	1022.6	1022.6	0.89	0.11
4	1022.5	1022.6	0.81	0.54
5	1022.6	1022.6	0.7	0.01

Na Tabela 2 estão apresentados os valores referentes à média dos 10% da base e dados de treino, a média dos 90% dos dados previstos e os seus respetivos desvios padrões. Como se pode observar existe uma grande diferença entre a média e o desvio padrão quando se trabalha apenas com 10% dos dados de treino em comparação com a média original dos cinco sensores (Tabela 1). Por outro lado, quando se comparam as médias e os desvios padrões referentes aos 10% dados treino e aos dados previstos não existe diferença significativa, isto ocorre, devido ao facto dos dados previstos serem em função dos dados treino, igualando assim os seus valores da base de treino do modelo. Seguidamente irá ser apresentada a tabela referente à utilização de 50% da base de dados de treino e será então verificado se existe alguma diferença significativa entre os dados reais e os previstos:

Tabela 3- Caracterização dos dados referentes a utilização de 50% de valores de treino

Sensor	Média	Média valores previstos	Desvio Padrão base de dados de treino	Desvio Padrão valores previstos
1	1022.6	1022.6	0.66	0.05
2	1022.5	1022.6	0.66	0.01
3	1022.6	1022.5	0.66	0.01
4	1022.6	1022.6	0.62	0.02
5	1022.7	1022.7	0.63	0.08

Comparando com os valores da Tabela 3 é possível verificar que a média não sofre praticamente nenhuma variação tendo em conta a diferença de tamanho da base de dados de treino utilizado, o desvio padrão sofre algumas variações, tendendo para zero. A nível de valores obtidos pode-se afirmar que neste caso em particular a diferenciação do tamanho da base de dados de treino utilizados para efetuar a previsão não produziu diferenças significativas, quer a nível da média quer a nível do desvio padrão. Tendo em conta que o alerta final é elaborado com base nestes valores será de esperar que, neste caso, não surjam qualquer tipo de diferenças a nível de alertas.

Resultados simulação I - Variação da constituição das bases de dados.

Uma vez que já foram obtidos os resultados para variação de tamanho da base de dados de treino, passar-se-á seguidamente ao estudo subsequente, variar a constituição das bases de dados de teste e de treino.

De salientar que foram efetuadas todas as variações de treino quer para os 10% de dados de treino quer para os 50%, de maneira a perceber qual seria a diferença provocada nos alertas. Porém tendo em conta a similaridade dos resultados irão ser apresentados apenas os referentes aos 50% de dados de treino.

Tendo em conta os resultados obtidos anteriormente este cenário produzirá o menor desvio padrão, logo resultará numa aproximação mais correta aos dados medidos pelos sensores, sendo os outros resultados apresentados no capítulo dos anexos.

Inicialmente serão expostos os gráficos referentes à variação de constituição número 1 (Fig. 5.9), que consiste na base de dados de treino com valores somente sem chama e base de dados de teste com valores somente na presença de chama:

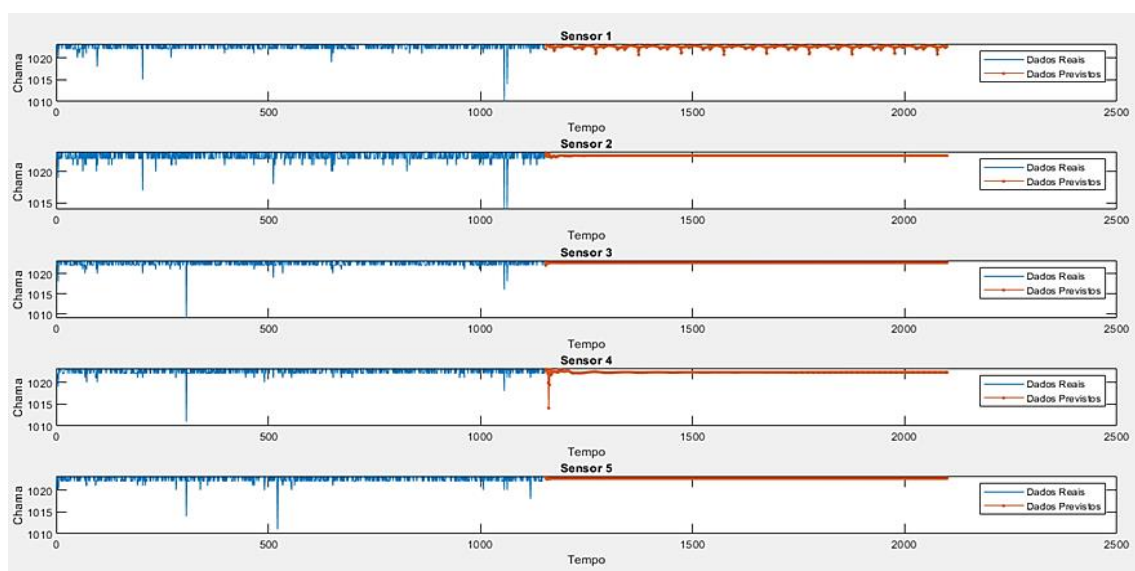


Figura 5.9- Variação de constituição número 1.

Na Figura 14 estão presentes as previsões dos cinco sensores. É possível observar que existem diferenças nas previsões entre os sensores e que os valores estão entre os esperados. O número de dados referente à base de dados de treino utilizados nesta variação foi de cerca de 1204 e cerca de 946 dados referentes a base de dados de teste.

Seguidamente foi testado o algoritmo considerando a variação número 2 (Fig. 5.10), constituída pela base de dados de treino com valores somente na presença de chama e base de dados de teste com valores somente sem chama:

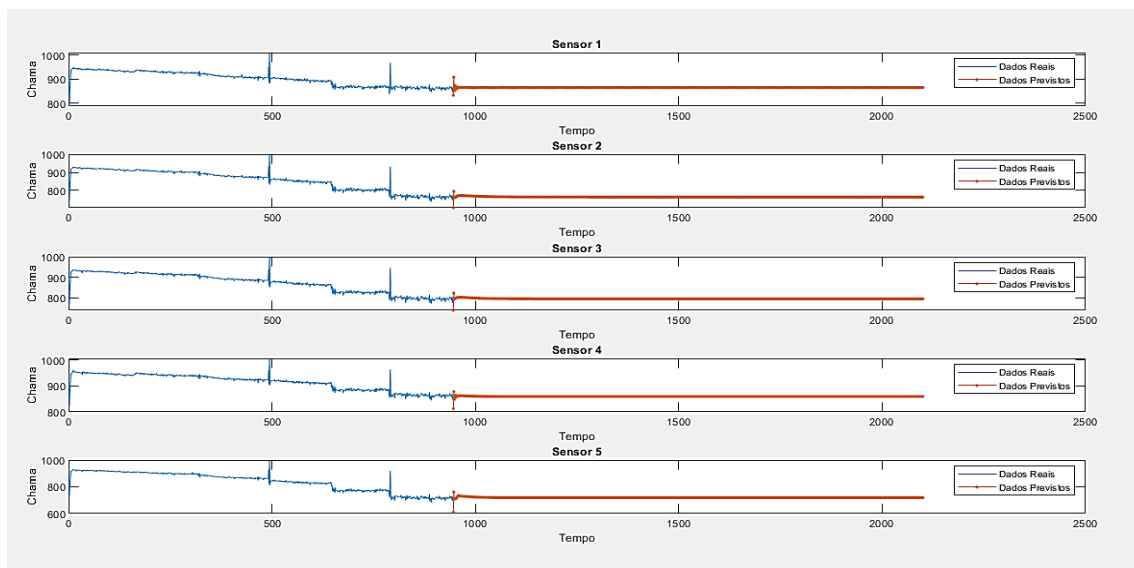


Figura 5.10- Variação de constituição número 2.

Neste conjunto de gráficos já é possível observar diferenças relativamente à Fig. 5.9. É possível notar que as previsões entre os sensores não apresentam diferenças significativas entre si, e os valores previstos estão entre 600 e os 800. Os valores obtidos são significativamente diferentes dos obtidos na Tabela 3 onde a base de dados de treino é constituída com valores somente sem chama e base de dados de teste com valores somente na presença de chama. Isto acontece devido ao facto de a base de dados de treino ser constituída com cerca de 946 valores todos referentes a situação de chama.

A variação número 3 (Fig. 5.8), tem por base a não diferenciação de dados com ou sem chama, isto é, a base de dados de treino e base de dados de teste são constituídas pela totalidade dos valores. Tendo em conta que os resultados referentes a esta variação já foram apresentados anteriormente na Fig. 5.8, não existe necessidade de apresentar os gráficos novamente. Relativamente a esta variação podemos afirmar que as previsões se mantêm estáveis num único valor, não sofrendo grandes variações. Esta variação foi elaborada com cerca de 2150 valores quer da base de dados de teste quer da base de dados de treino.

Por fim, serão apresentados os resultados provenientes da variação número 4 (Fig. 5.11), esta previsão foi elaborada recorrendo a um conjunto de 1892 referente à base de dados de treino sem qualquer presença de chama e 1892 alusivos à base de dados de teste onde a chama já estava acesa, resultando na seguinte Figura:

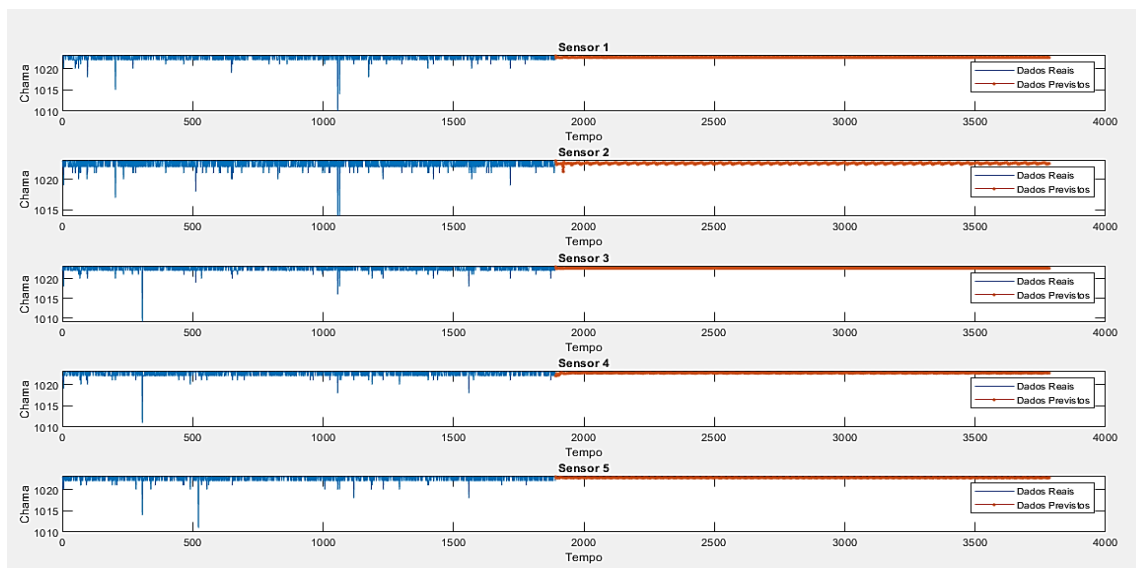


Figura 5.11- Variação de constituição número 4.

Como se pode observar na Fig. 5.11, existem poucas diferenças visuais quando comparados com a variação número 3 (Fig. 5.8), sendo que todos os sensores à exceção do sensor dois se mantêm constantes num único valor. Não existindo assim qualquer vantagem no uso desta estratégia em relação a variação número 3.

Resultados simulação I - Variação do método de identificação dos alertas.

Uma vez que os resultados referentes às variações do tamanho e da constituição das bases de dados foram já apresentados, seguidamente serão apresentados os resultados referentes à variação do método de identificação dos alertas.

Os resultados referentes aos alertas são provenientes de dois diferentes métodos de identificação, como se demonstrou no Capítulo 4. Assim sendo, inicialmente serão apresentados os resultados graficamente e por fim serão apresentados em forma de tabela para uma melhor compreensão.

Os resultados aqui presentes graficamente são provenientes da variação número 3, base de dados de treino e base de dados de teste com todos os valores e com a variação de tamanho de 10% da base de dados de treino, pois visualmente são mais intuitivos, sendo que todos os outros resultados serão apresentados nos anexos.

Inicialmente irá ser apresentada a imagem referente à primeira estratégia, que utiliza a diferença entre a base de dados de teste e os valores previstos:

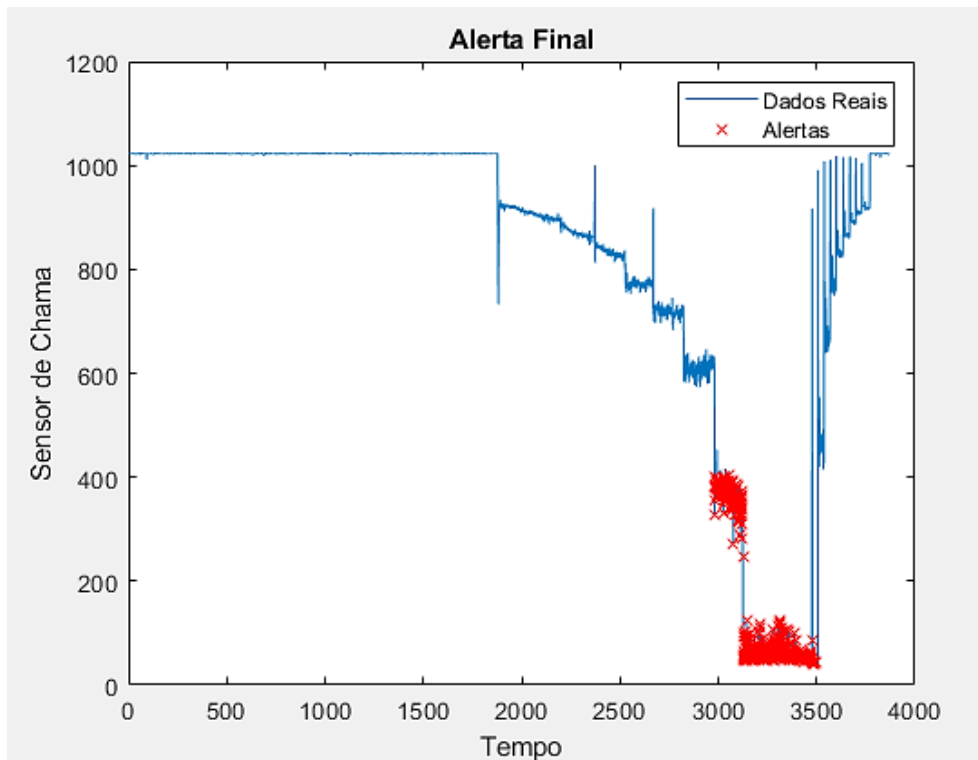


Figura 5.12- Alerta final primeira estratégia.

Como se pode verificar esta estratégia não produz os resultados esperados, uma vez que apenas apresenta alertas quando a chama está a poucos centímetros do sensor, mais especificamente, o primeiro alerta é dado quando a chama se encontra a cerca de 30 cm do sensor.

A segunda opção funcionava de um modo mais simples, comparando diretamente os valores da base de dados de teste com a média menos o desvio padrão dos valores previstos resultando na seguinte imagem:

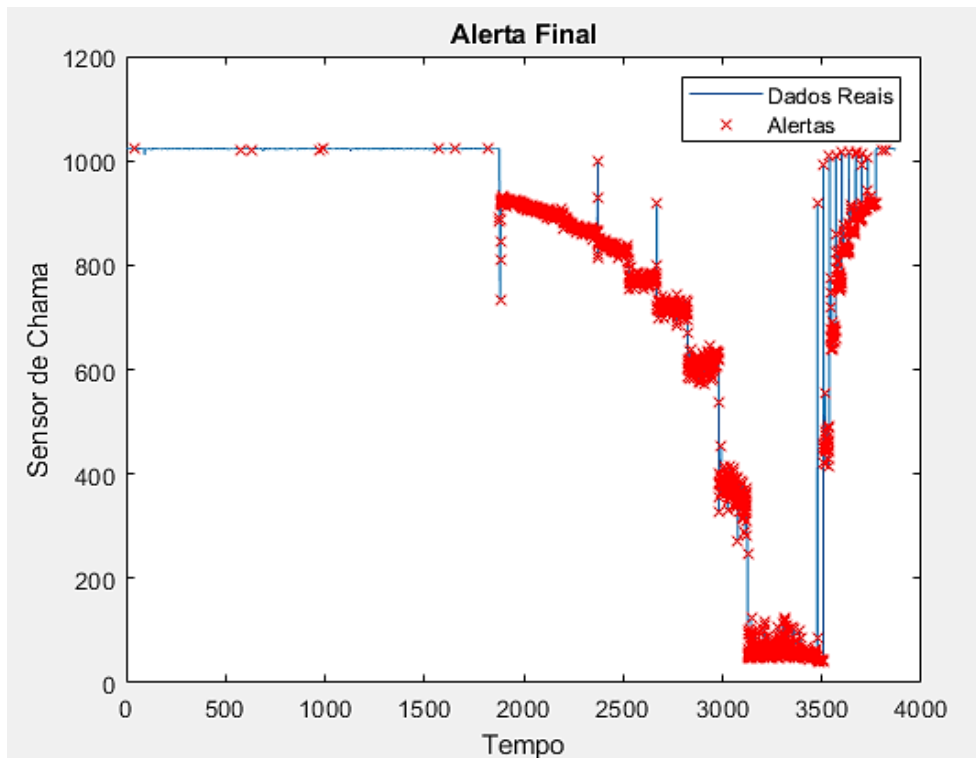


Figura 5.13- Alerta final segunda estratégia.

Neste caso, os alertas foram obtidos adequadamente pois no preciso momento que a chama é acesa o sistema começa a alertar de forma contínua, porém, por outro lado, esta estratégia produz também alguns falsos alertas o que nos sugere que será necessário uma maior quantidade de sensores ou de dados para corrigir este problema.

De maneira a ilustrar melhor os resultados referentes às diferentes estratégias serão apresentadas seguidamente as tabelas contendo os valores referentes aos alertas:

Tabela 4- Resumo dos alertas primeira estratégia

Sensor	Nº de alertas individuais	Nº de alertas final	Nº ideal de alertas final
1	702	497	1892
2	693		
3	529		
4	553		
5	554		

Como se pode verificar na tabela 4, em concordância com o que já se tinha observado no gráfico 7 a primeira estratégia apresenta um baixo número de alertas quando

comparados com o número de alertas ideal, número de instantes que a chama esteve acesa. Resultando assim numa percentagem de acerto de cerca de apenas 26.3%.

Seguidamente será apresentada a tabela referente à segunda estratégia de elaboração de alertas:

Tabela 5- Resumo dos alertas segunda estratégia

Sensor	Nº de alertas individuais	Nº de alertas final	Nº ideal de alertas final
1	1936	1903	1892
2	1973		
3	2639		
4	2572		
5	2500		

Neste caso, contrariamente à primeira estratégia, existe um número de alertas superior ao pretendido, esta estratégia de resolução apresenta cerca de 11 falsos alertas quando comparada com o número ideal de avisos. Tendo em conta a percentagem de acerto, é possível que aliando estes sensores de chama com outro tipo de sensores, nomeadamente temperatura e humidade se consiga evitar estes falsos alertas.

5.2. Simulação II

Os dados seguidamente apresentados, foram obtidos através dos cinco sensores de chama na presença de uma pequena queimada, controlada, como se ilustra na Figura 5.2.

5.2.1. Caracterização dos dados

A segunda simulação foi efetuada ao ar livre num dia com bastante luminosidade, dificultando assim a deteção de incêndio, os sensores foram colocados a três metros da ignição resultando nos seguintes valores:

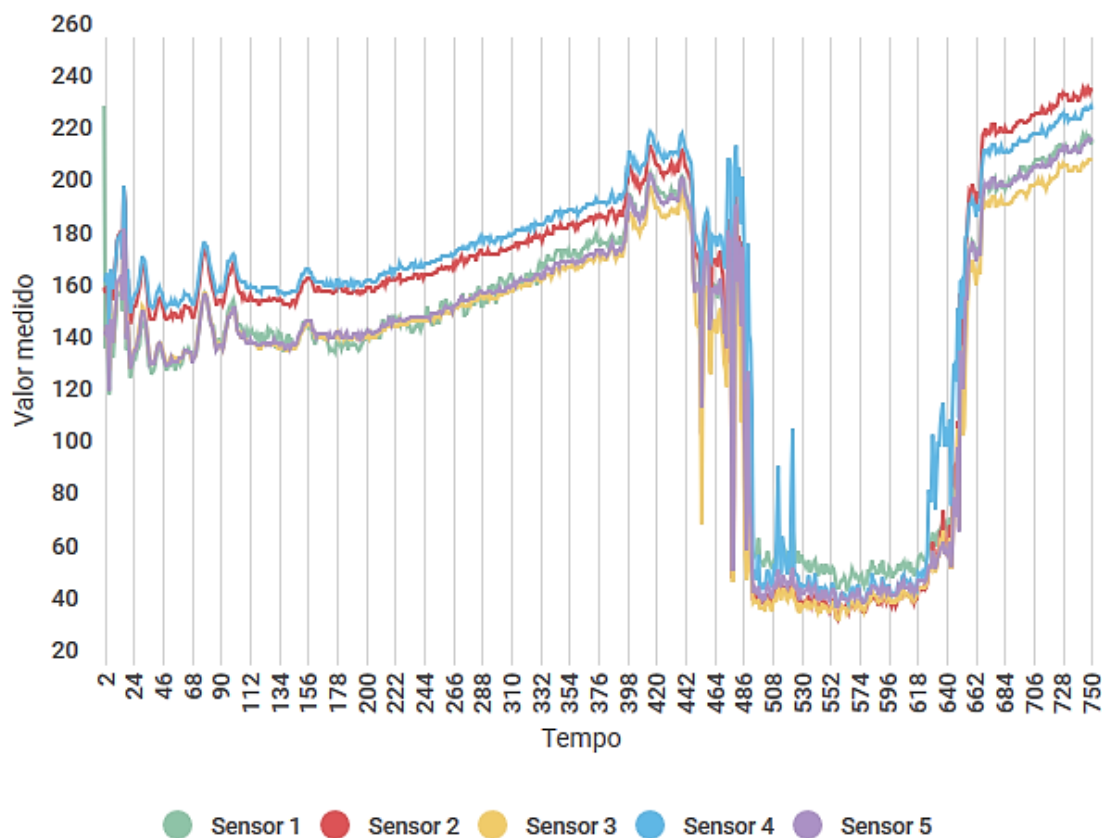


Figura 5.14- Dados referentes a segunda simulação.

O gráfico é referente a um conjunto de cerca de 750 unidades de tempo, sendo que cada unidade corresponde a 2 segundos perfazendo um total de 25 minutos. O protocolo, neste caso está dividido da seguinte maneira:

- Do instante 0 até ao instante 450 não existiu qualquer tipo de ignição servindo assim de teste de controlo;
- No instante 450 foi acesa uma pequena chama, que demorou alguns instantes até se propagar como se pode observar no gráfico e permaneceu acesa até ao instante 650;
- No instante 650 a chama foi extinta, não existindo qualquer ignição até ao final da simulação.

Devido ao facto de a luz solar estar em decaimento, é possível observar que os dados obtidos através dos sensores tendem a subir em função do tempo. Neste caso é possível observar a influência da luz neste tipo de sensor começando o teste com valores de 140 e terminando o mesmo 25 minutos depois com valores de 220.

Seguidamente será apresentada a tabela referente à média e ao desvio padrão dos dados:

Tabela 6- Caracterização dos dados da segunda simulação

Sensor	Média	Desvio Padrão
1	144.09	49.86
2	153.47	59.98
3	138.24	52.46
4	159.33	55.87
5	141.74	52.66

As médias referentes aos cinco sensores variam entre [138.24, 159.33]. Nestes dados pode-se comprovar a influência da luminosidade nos sensores, comparando com a Tabela 1 referente à primeira simulação onde os valores das médias variavam entre [846.51, 953.78]. Verificando-se assim a diferença de valores nas diferentes fases do dia podendo originar problemas quer de falsos alertas como também de ignições não identificadas devido à excessiva luminosidade nos sensores. Uma forma de evitar esta situação seria introduzir outro tipo de sensores como por exemplo sensores de temperatura e humidade.

5.2.2. Resultados obtidos

Tendo em conta os resultados obtidos na simulação I, foi possível observar que a variação da constituição da base de dados de treino não surtiu nenhuma melhoria em relação à constituição utilizada por defeito (variação 3) onde todos os dados são usados quer para a base de dados treino quer para a base de dados de teste. Devido a isto, na segunda simulação apenas foi variada a quantidade de dados de treino utilizada, sendo apresentadas seguidamente as tabelas referentes a 10% e 50% de dados de treino respetivamente e os seus respetivos gráficos de alerta.

Tabela 7- Caracterização 10% de dados de treino segunda simulação

Sensores	Média da base de dados de treino	Média Valores Previstos	Desvio Padrão da base de dados de treino	Desvio Padrão Valores Previstos	Nº de alertas
1	141.74	141.66	10.16	10.57	167
2	161.15	161.41	10.77	10.70	170
3	144.78	144.49	10.71	9.87	177
4	164.82	163.26	9.49	7.49	165
5	144.12	147.05	11.25	11.68	168

Observando a Tabela 7 é possível perceber que os dados obtidos tendem a seguir o padrão da primeira simulação, isto é, nota-se similaridade entre a média total dos sensores quando comparada com 10% da base de dados de treino, o que nos permite afirmar que os dados usados são consistentes, não existindo discrepâncias significantes entre eles.

O número ideal de alertas finais neste caso era de 200, obtendo-se 164, alcançando assim uma taxa de acerto de 82%. Os resultados obtidos na simulação II revelaram-se um pouco desfasados dos esperados, devido ao facto de a chama se ter praticamente extinguido após a sua ignição, não existindo chama durante alguns segundo resultando assim numa quebra de padrão, observável na seguinte figura no intervalo de tempo entre [392, 408]:

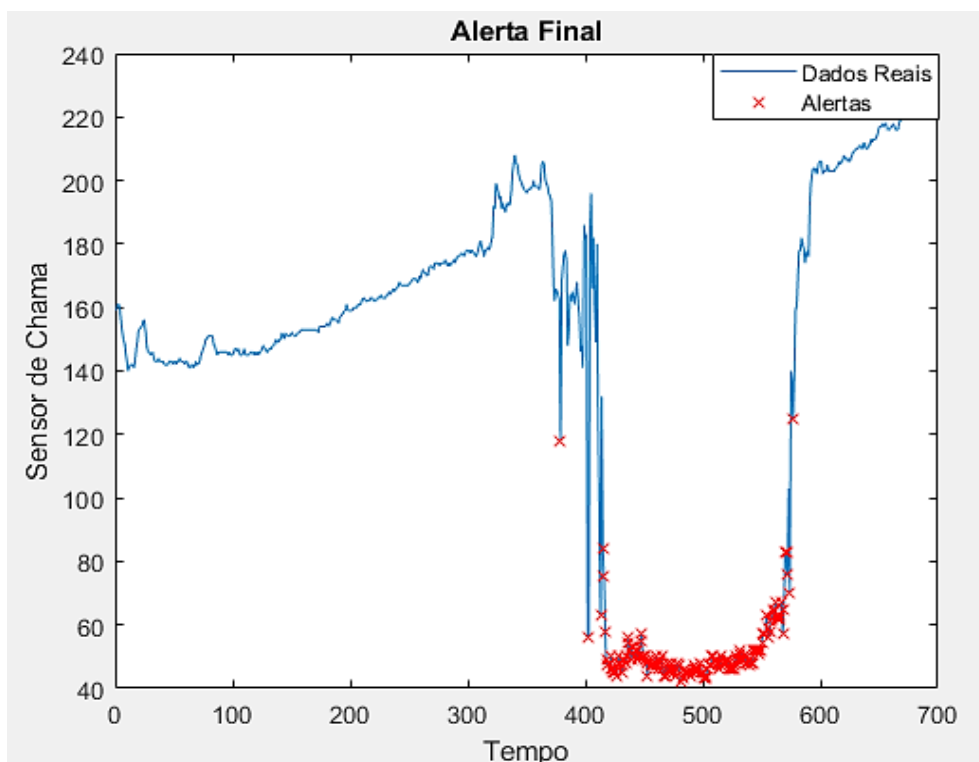


Figura 5.15- Alerta final da simulação II, 10% da base de dados de treino.

Seguidamente foi aplicado o algoritmo com 50% de dados de treino para assim comparar qual seria a melhor opção, obtendo a seguinte tabela resumo:

Tabela 8- Caracterização 50% de dados de treino segunda simulação

Sensores	Média da base de dados de treino	Média Valores Previstos	Desvio Padrão da base de dados de treino	Desvio Padrão Valores Previstos	Nº de alertas
1	152.78	176.91	12.93	6.19	197
2	168.98	185.74	10.67	3.49	197
3	152.06	172.16	10.97	5.11	205
4	173.12	173.63	11.00	7.86	166
5	152.94	172.73	11.65	4.14	197

Como se pode verificar, neste caso as médias obtidas através do uso de 50% da base de dados de treino [152.06, 173.12], ultrapassam no geral os valores das médias referentes a totalidade dos dados [138.24, 159.33], tornando-se assim uma amostra um pouco sobredimensionada. Porém, esta diferença de média não se refletiu de todo no resultado, pois neste caso, foi conseguida uma percentagem de acerto de 83%, obtendo 166 alertas dos 200 ideais. Este aumento deveu-se ao facto de a quantidade de alertas individuais ser superior ao caso anterior. Como é possível observar na figura seguinte:

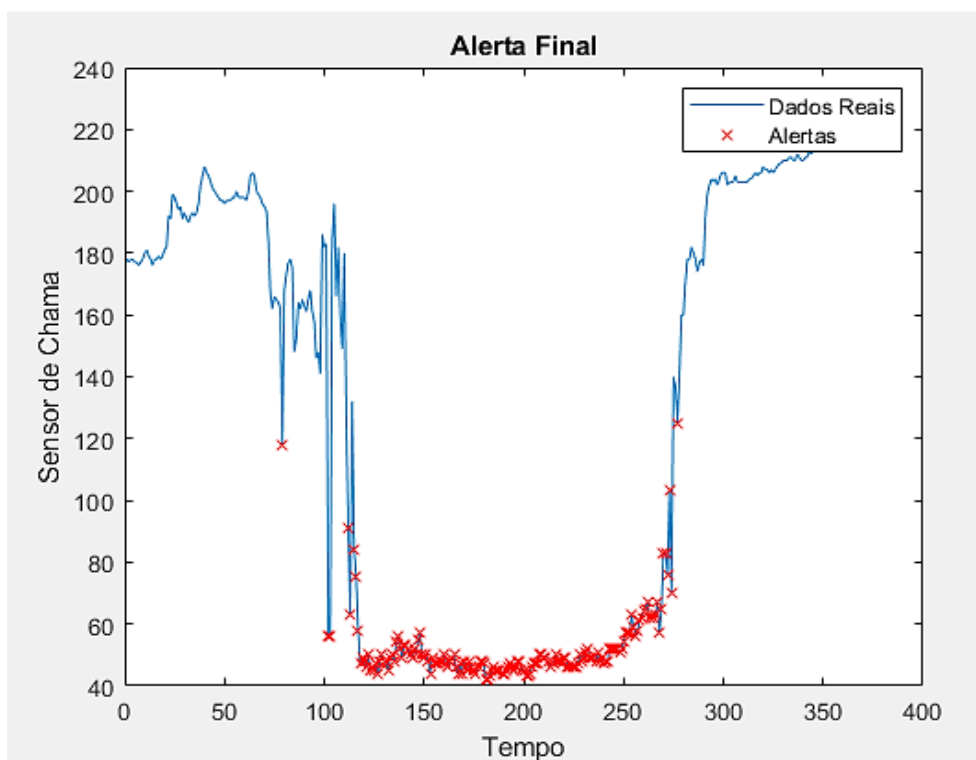


Figura 5.16- Alerta final da simulação II, 50% da base de dados de treino.

De salientar que ambos os alertas foram obtidos de acordo com a segunda estratégia uma vez que esta se demonstrou mais eficaz na primeira simulação. Os gráficos referentes às previsões estão disponíveis nos anexos.

5.3. Simulação III

Este conjunto de dados, é proveniente de um dos módulos sensorizados piloto apresentados na Secção 4.2. - Sensores utilizados, instalado neste momento no campus do IPB.

O módulo sensorizado possui três tipos de sensores: cinco sensores de chama, um sensor de medição de temperatura e ainda um sensor de medição de humidade relativa do ar.

5.3.1. Caracterização dos dados

Os dados são referentes às datas entre 21 de julho de 2020 e 18 de agosto do mesmo ano, durante este período não ocorreu qualquer tipo de ignição sendo, portanto apenas dados de controlo. No período referido foram obtidas informações a cada minuto constituindo esta a nossa unidade de tempo. Durante o período de estudo ocorreram diversas mudanças de temperatura e de humidade como é de esperar ao longo dos dias, variando os dados referentes a temperatura entre os 21° e os 38° celsius e os valores da humidade do ar alteraram entre os 11% e os 95%.

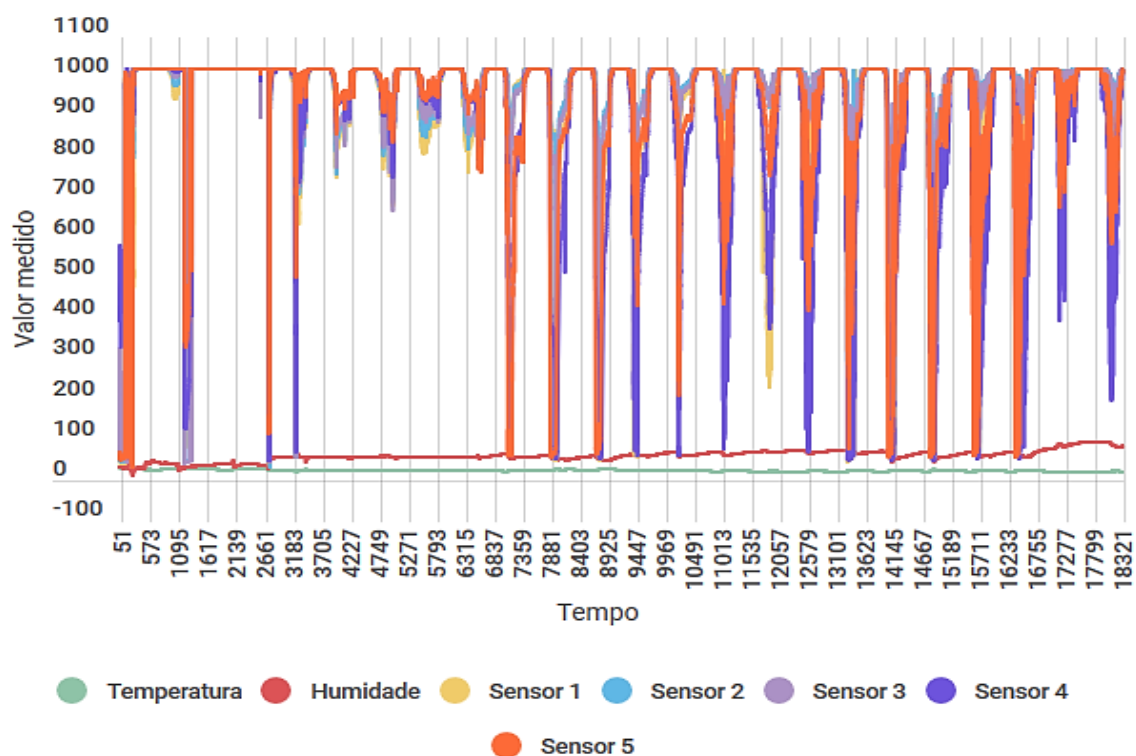


Figura 5.17- Dados referentes a terceira simulação.

Como se pode observar é bastante notável a diferença nos valores referentes aos sensores de chama de acordo com o período do dia em questão, é notável também as suas variações mesmo durante o dia em períodos mais ou menos nublados. Os valores referentes à temperatura e à humidade mantêm-se dentro do esperado aumentado a humidade nos períodos noturnos e diminuindo a temperatura, funcionando de maneira oposta nos períodos de maior calor durante o dia. Seguidamente será exposta uma tabela de modo a apresentar a média e o desvio padrão de cada sensor:

Tabela 9- Caracterização dos dados da terceira simulação

Sensor		Média	Desvio Padrão
Sensores de Chama	1	941.32	187.84
	2	970.42	148.76
	3	969.42	143.09
	4	895.15	232.65
	5	925.14	209.26
Temperatura em °C		24.24	2.01
Humidade relativa		60.45	13.40

Como se pode observar, ambos os cinco sensores de chama apresentam dados relativamente similares entre si, não existindo nenhuma discrepância entre eles, a temperatura e a humidade relativa apresentam valores normais para a época em questão, de salientar que o conjunto total de dados desta simulação é constituído por 18321 entradas.

5.3.2. Resultados obtidos

Em semelhança com a Simulação II, neste caso irá ser também variado o tamanho da base de dados de treino, o processo de treino será constante e irá ser o processo de treino e confirmação com todos os valores (variação 3).

Inicialmente serão apresentadas as tabelas referentes a cada uma das variações de tamanho da base de dados de treino, as figuras relativas aos alertas não serão apresentadas devido ao facto de o número de alertas esperados neste caso ser de zero, sendo apresentadas somente nos anexos.

A primeira tabela apresentada é referente a utilização de 10% de dados de treino e 90% da base de dados de teste:

Tabela 10- Caracterização 10% de dados de treino terceira simulação

Sensores		Média da base de dados de treino	Média Valores Previstos	Desvio Padrão da base de dados de treino	Desvio Padrão Valores Previstos	Nº de alertas
Sensores de Chama	1	876.90	868.23	330.80	21.51	2041
	2	876.65	216.11	340.07	4.47	118
	3	882.26	1028.3	330.73	54.71	4283
	4	922.36	1220.6	254.88	47.59	6592
	5	894.92	1009.6	305.75	76.06	4276
Temperatura em °C		27.67	29.59	1.69	0.09	17
Humidade relativa		36.50	37.96	6.62	0.24	264

Como se pode observar a média e o desvio padrão dos 10% de dados de treino, reproduzem uma boa amostra em relação aos dados dos sensores. Por outro lado, a média dos dados previstos possui algumas variâncias de valores, como por exemplo no Sensor 2. Tendo em conta que os alertas são obtidos através da média menos o desvio padrão, faz sentido que neste caso o Sensor 2 seja o sensor que mais se aproxima da realidade, à parte do sensor da temperatura, podendo confirmar estes dados mais intuitivamente com as imagens presentes nos anexos. Relativamente aos resultados podemos expor que foram

satisfatórios, uma vez que, tendo em conta o elevado número de alertas individuais seria expectável um maior número de alertas finais. Porém tal não se verificou obtendo no total zero alertas que coincide com o número ideal de alertas uma vez que não existiu nenhuma ignição.

Calculados os alertas para os 10% de dados de treino, passar-se-á de seguida ao cálculo utilizando 50% da base de dados de treino, apresentando os restantes na seguinte tabela:

Tabela 11- Caracterização 50% de dados de treino terceira simulação

Sensores		Média da base de dados de treino	Média Valores Previstos	Desvio Padrão da base de dados de treino	Desvio Padrão Valores Previstos	Nº de alertas
Sensores de Chama	1	944.79	928.72	193.65	46.36	1671
	2	951.74	573.40	190.27	164.17	50
	3	952.63	615.87	183.92	30.23	72
	4	933.31	350.26	207.98	11.16	590
	5	933.19	1037.2	228.39	60.64	3766
Temperatura em °C		25.56	24.81	1.76	0.24	990
Humidade relativa		51.41	22.38	9.91	1.91	0

Contrariamente ao que aconteceu na simulação I, neste caso a utilização de maior quantidade de dados de treino para obter a previsão já demonstrou diferenças em relação a utilização de apenas 10%. Como se pode observar as médias da base de dados de treino estão ainda mais próximas das médias da base de dados inicial, os dados previstos possuem também algumas variações, em similaridade com a tabela anterior o Sensor 2 é o sensor de chama que mais se aproxima da quantidade de alertas pretendida, sendo que o sensor referente a humidade relativa atinge com sucesso o número de alertas pretendidos. Mais uma vez comprovou-se a eficácia do algoritmo uma vez que o número de alertas final é concordante com o número pretendido.

Capítulo 6 - Conclusões e trabalhos futuros

Neste capítulo serão apresentadas as conclusões deste trabalho e ainda sugestões de trabalho futuros na continuação deste projeto.

6.1. Conclusões

As conclusões retiradas deste estudo podem ser divididas em dois temas, as conclusões relativas aos sensores e ao seu comportamento e as conclusões referentes ao algoritmo e ao seu funcionamento.

A estratégia de utilização de sensores para vigilância e recolha de dados possui, como foi possível observar, inúmeras vantagens. Começando por referir o seu baixo custo quando comparados com os outros métodos disponíveis neste momento, o tamanho e facilidade de instalação destes sensores é um grande ponto a seu favor, podemos adicionar à lista das vantagens a duração da sua bateria e a facilidade de transmissão dos dados uma vez que apenas se lida com valores, não existindo a necessidade de transmitir imagens. O comportamento dos sensores de chama revelou-se bastante satisfatório num ambiente controlado, isto é, quando é possível controlar a incidência solar sobre as lentes. Neste caso, obtêm-se resultados satisfatórios com relativamente poucos falsos alertas como foi possível observar na primeira simulação. Porém, como se conseguiu comprovar com a terceira simulação, se não existir um controlo da luz solar, a utilização deste tipo de sensores isoladamente, apresenta uma taxa de falsos alertas preocupante. Isto ocorre devido às variações da intensidade da luz solar durante o dia, obtendo-se valores referentes aos sensores de chama com um intervalo de 29 a 1023 o que nos torna praticamente impossível conseguir diferenciar uma ignição de um raio de sol mais intenso. A utilização simultânea de todos os sensores vem resolver este problema uma vez que aliando a temperatura e a humidade aos sensores de chama consegue-se acabar com todos os falsos alertas, lembrando ainda que com a adição do sensor ultravioleta conseguir-se-á ampliar ainda mais estes bons resultados.

Relativamente ao algoritmo, foram elaboradas várias variações de maneira a encontrar a melhor combinação possível, pode-se referir que quando se trabalha com um pequeno conjunto de dados, como aconteceu com a primeira e a segunda simulação, não existe praticamente nenhuma diferença relativamente à utilização de 10% ou 50% de dados reais para efetuar a previsão. O mesmo não acontece quando se trabalha com maiores conjuntos de dados como aconteceu na última simulação, foi possível observar que os valores referentes aos 50% se aproximavam mais do conjunto inicial como já seria de esperar. Os processos de treino também revelaram pouca significância na primeira simulação, aliando isto ao facto de ser a forma mais fácil para a obtenção da previsão, ficou predefinido que o treino seria sempre efetuado com todos os valores provenientes dos sensores sem qualquer pré divisão. A última variação foi a obtenção dos alertas, neste caso ficou em clara vantagem a segunda estratégia produzindo quer em laboratório quer em contexto real melhores resultados. Pode-se então concluir que o algoritmo cumpre o seu propósito, porém pode ainda sofrer algumas melhorias a nível de rapidez de processamento e melhoramento de falsos alertas.

6.2. Trabalhos futuros

No futuro pretende-se melhorar ainda mais o software, para tal algumas alterações irão ser efetuadas, nomeadamente a implementação dos sensores em falta, sensor ultravioleta e o sensor de humidade do solo. Seguidamente serão estudadas, tendo em conta a entrada de mais dois sensores no sistema, as variações de treino do algoritmo de maneira a torná-lo mais eficiente, reduzindo o número de alertas referentes aos sensores de chama.

Um passo importante será uma mudança de design do software, uma vez que o utilizado neste momento é pouco intuitivo tornando-se um pouco confuso para o utilizador. À parte do design, é ainda almejado o teste de outros algoritmos de maneira a comparar taxas de acerto e rapidez de execução.

Num período mais alargado de tempo pretende-se ampliar o número de módulos com que o sistema trabalha, isto é, ajustar o algoritmo a receber dados de conjuntos de módulos sensorizados. Uma vez que o algoritmo esteja devidamente adaptado, é pretendido aplicar o *software* no contexto real, isto é, receber os dados diretamente dos sensores e fazer o processamento instantaneamente dos mesmos, avaliando a situação e

enviando os alertas para as entidades competentes, obtendo assim um *software* mais fiável dia para dia uma vez que a base de dados seria aumentada a cada minuto tornando assim as previsões cada vez mais corretas.

Referências Bibliográficas

- Riahi, Y., 2018. Big Data and Big Data Analytics: Concepts, Types and Technologies. Outubro, 5(524-528), pp. 524-528.
- Alkhatib, A., 2014. A Review on Forest Fire Detection Techniques. *International Journal of Distributed Sensor Networks*, 5 Março, Issue 1-6.
- Alves, J. M., 2018. *Deteção de incêndios florestais com recurso a Deep Learning e Visão Computacional*, s.l.: s.n.
- Alves, R. A. V., 2018. *Emprego de meios tecnologicos na vigilância florestal em Portugal*, s.l.: IUM.
- Bengio, Y., Courville, A. & Vincent, P., 2013. Representation learning: A review and new perspectives.. 35(1798-1803).
- Bezerra, E., 2016. *Tópicos em Gerenciamento de Dados e Informações*. s.l.:SBC.
- Brito, T. et al., 2020. *Optimal Sensors Positioning to Detect Forest Fire Ignitions*. s.l., s.n.
- Calle, A., Casanova, J. L. & Romo, A., 2006. Fire detection and monitoring using MSG Spinning Enhanced Visible and Infrared Imager (SEVIRI) data. *Journal of Geophysical Research Atmospheres*, 7 Setembro, 111(3).
- Changyong, C. et al., 2014. Early On-Orbit Performance of the Visible Infrared Imaging Radiometer Suite Onboard the Suomi National Polar-Orbiting Partnership (S-NPP) Satellite. *IEEE Transactions on Geoscience and Remote Sensing*, 2 Fevereiro.52(1142-1156).
- de Dios, J. R. M. et al., 2007. Computer vision techniques for forest fire perception. *Image and Vision Computing*, 9 Julho, 26(550-562).
- de Figueiredo, F. A. P., 2020. *Introdução à Inteligência Artificial e Machine Learning*, s.l.: s.n.
- EUFOFINET, 2012. *Project Detection Synthesis of Good Practices*, s.l.: s.n.
- Félix, F. & Lourenço, L., 2017. O tempo de resposta do ataque inicial a incêndios florestais nos espaços mais sensíveis de Portugal: o exemplo prático da serra da Lousã. *Territorium*, Issue 187-211.
- Goodfellow, I., Bengio, Y. & Courville, A., 2020. *Deep learning book*. [Online] Available at: <http://www.deeplearningbook.com.br>

- Governo, P., 2017. Despacho nº 9081-E/2017. *Diário da República - Gabinete do Secretário de Estado das Florestas e do Desenvolvimento*, 13 Outubro, Issue 1.
- Gusmão, J. M. N., 2017. *Aprendizagem Profunda: Estudo e Aplicações*, Évora: s.n.
- ICNF, 2017. [Online]
Available at: <http://www2.icnf.pt/portal/florestas/dfci/Resource/doc/rel/2017/10-rel-prov-1jan-31out-2017.pdf>
[Acedido em 21 Julho 2020].
- Kumar, V., 2018. Deep Learning as a Frontier of Machine Learning: A Review. *International Journal of Computer Applications*, 17, 182(22-30).
- Mathworks, 2020. [Online]
Available at: <https://www.mathworks.com/help/deeplearning/ug/time-series-forecasting-using-deep-learning.html>
[Acedido em 27 Julho 2020].
- NASA, 2020. *MODIS*. [Online]
Available at: <https://modis.gsfc.nasa.gov/about/specifications.php>
- Pellucci, P. R. S., de Paula, R. R., Silva, W. B. d. O. & Ladeira, A. P., 2011. Use of machine learning techniques in recognition of portuguese named entities. Em: Belo Horizonte: e-xacta.
- Pereira, D., Mendes, J. & Pereira, M., 2019. *Sensors*, Bragança: s.n.
- Rigo, D., Libertà, G. & Durrant, T., 2017. *Forest fire danger extremes in Europe under climate change: variability and uncertainty*, Luxembourg: Publications Office of the European Union.
- Rosa, J. L. G., 2011. *Fundamentos inteligência artificial*. Rio de Janeiro: GEN-Grupo Editorial Nacional/ LTC.
- Traiche, M. et al., 2008. Lidar Technique for Early Forest Fire Detection: Design and Development Aspects.. 1047(50-58), pp. 50-28.
- Utkin, A. B. et al., 2003. Feasibility of forest-fire smoke detection using lidar. *International Journal of Wildland Fire*, 12(159).
- Verma, N. & Singh, D., 2020. Analysis of cost-effective sensors: Data Fusion approach used for Forest Fire Application. *Materials Today: Proceedings*, 24(2283-2289).
- Wilson, C. & Davis, J., 1988. *Forest Fire Laboratory at Riverside and fire research in California: past, present, and future.*, California: General Technical Report (GTR).
- Yuan, C., Liu, Z. & Zhang, Y., 2016. *Vision-based forest fire detection in aerial images for firefighting using UAVs*. Arlington, IEEE.

Yuan, C., Liu, Z. & Zhang, Y., 2017. Fire detection using infrared images for UAV-based forest fire surveillance. Em: *2017 International Conference on Unmanned Aircraft Systems (ICUAS)*. Miami, FL, USA: s.n., pp. 567-572.

Yuan, C., Zhixiang, L. & Zhang., Y., 2015. *UAV-based forest fire detection and tracking using image processing techniques*. Denver, IEEE, pp. 639-643.

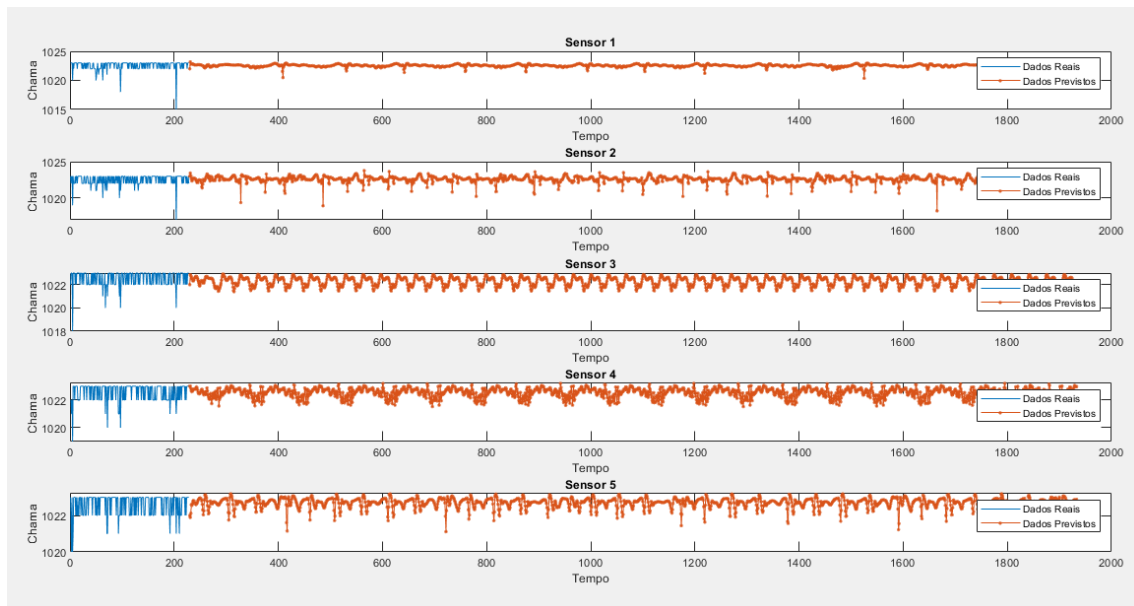
Anexos

Simulação I

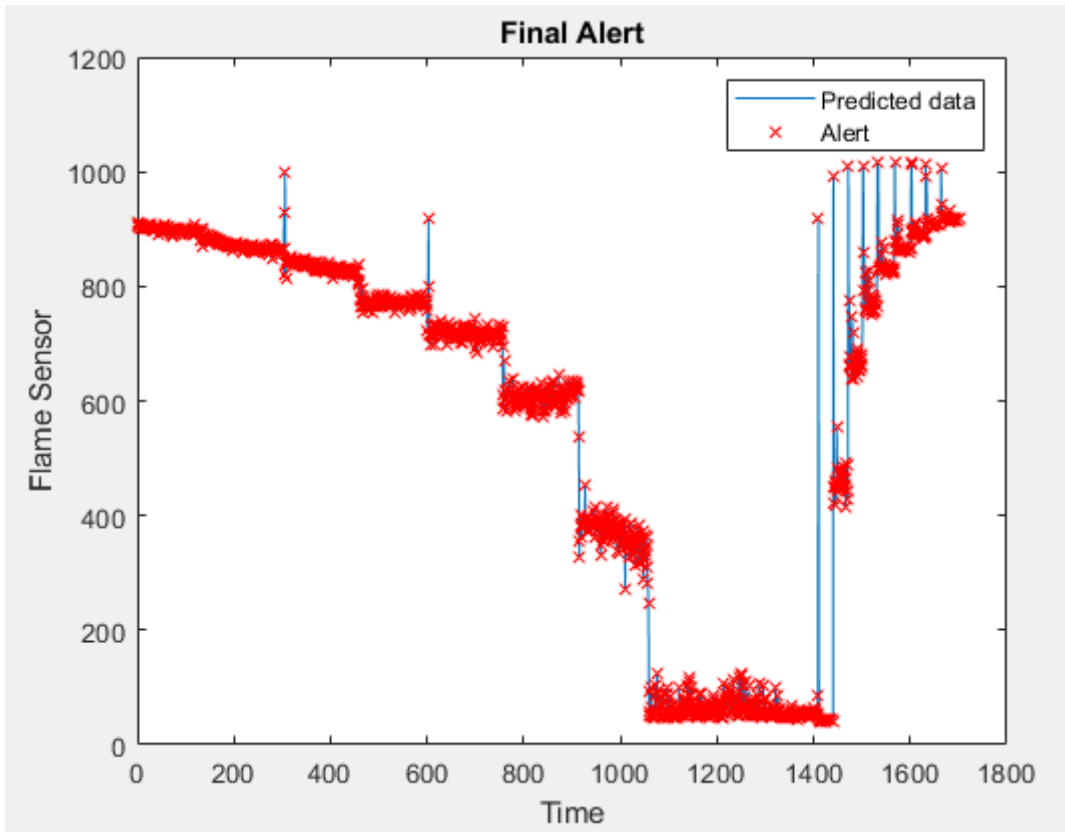
Gráficos referentes a utilização de 10% da base de dados treino:

Treino sem chama e confirmação com chama (Variação 1):

-Previsão:

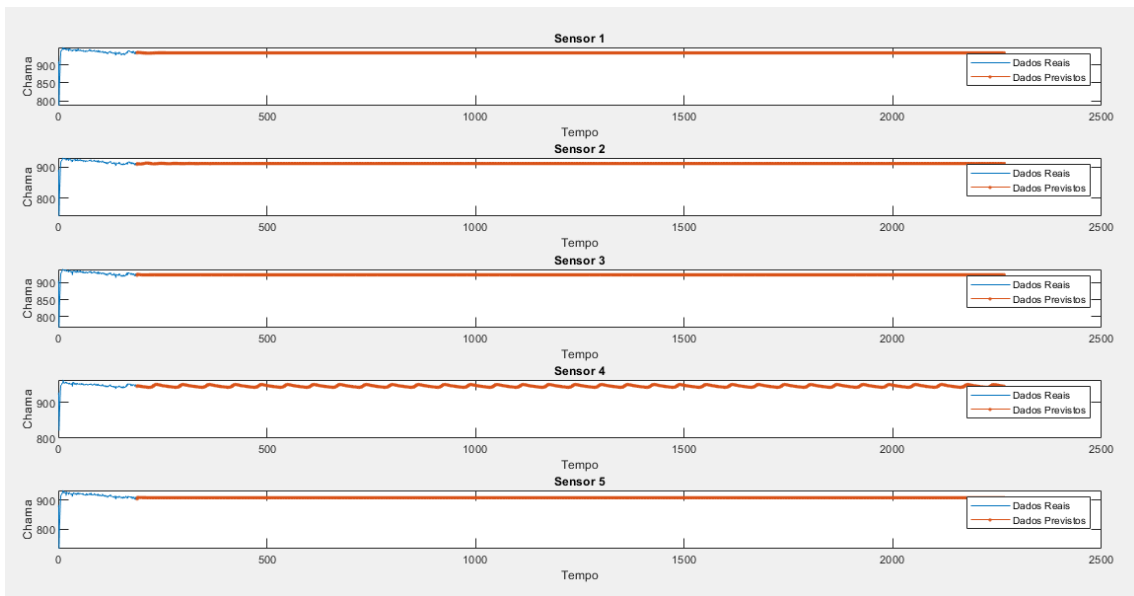


-Alerta final:

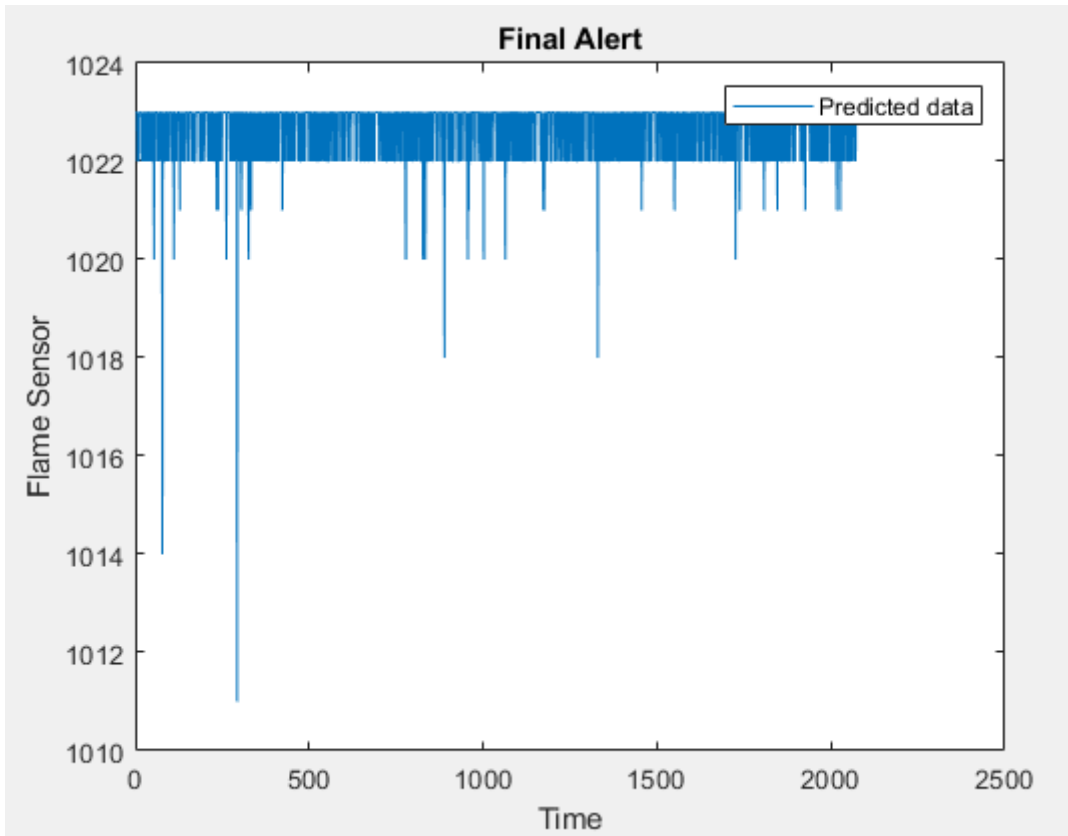


Treino com chama e confirmação sem chama (Variação 2):

-Previsão:

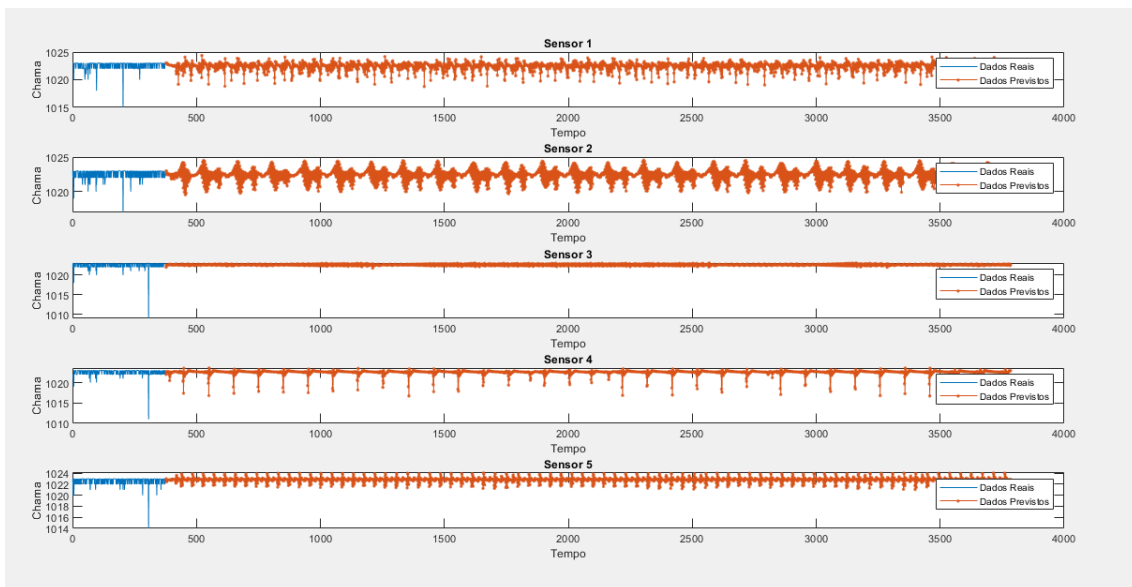


-Alerta final:

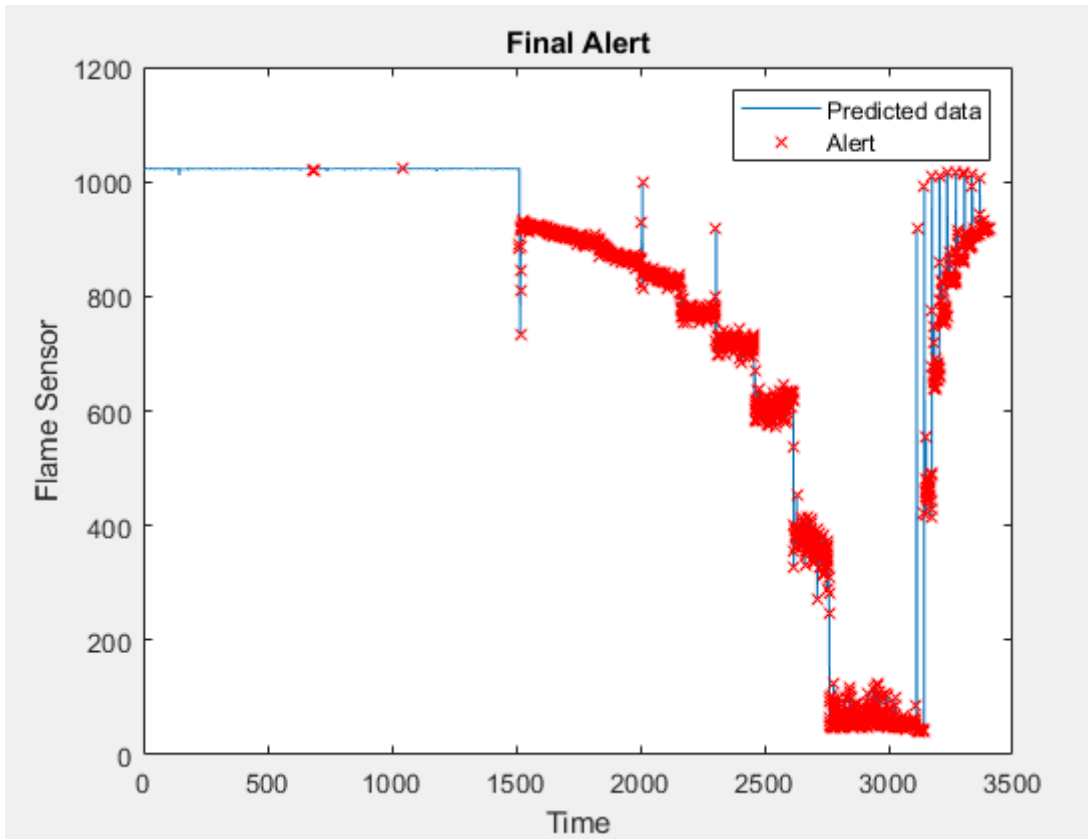


Treino com 50% sem chama e 50% com chama (Variação 4):

-Previsão:



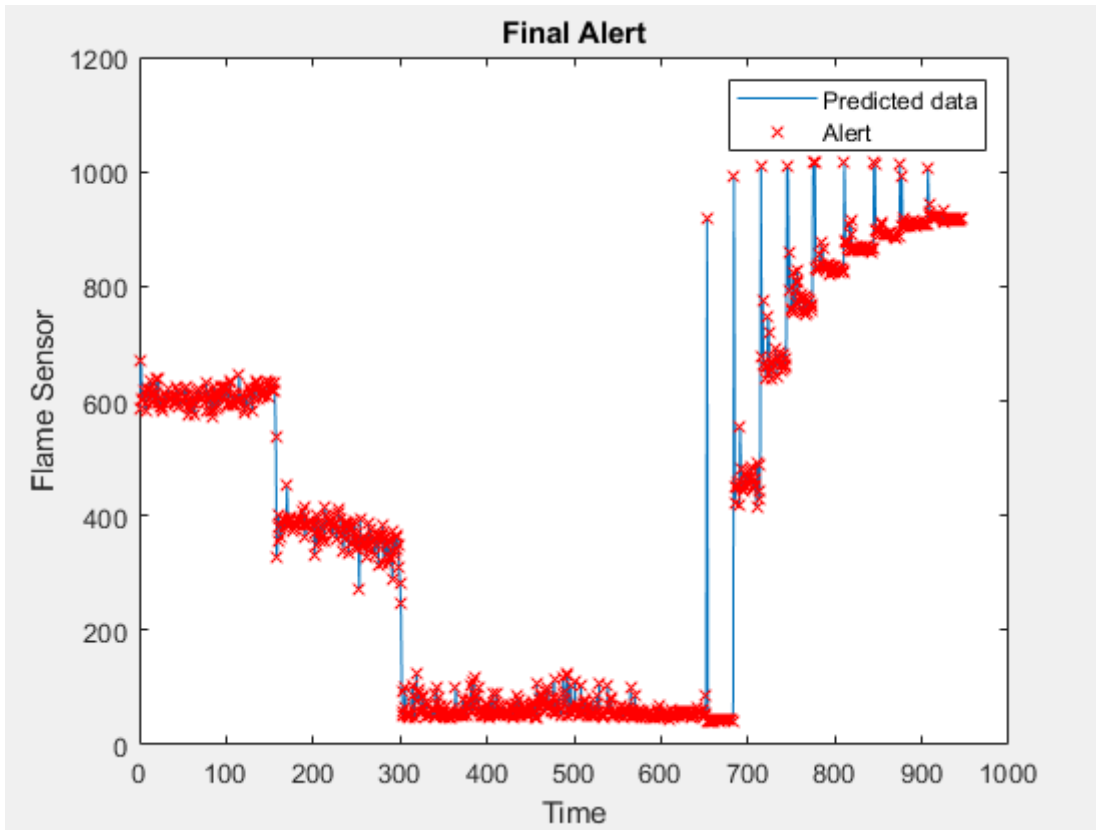
-Alerta final:



Gráficos referentes a utilização de 50% da base de dados de treino:

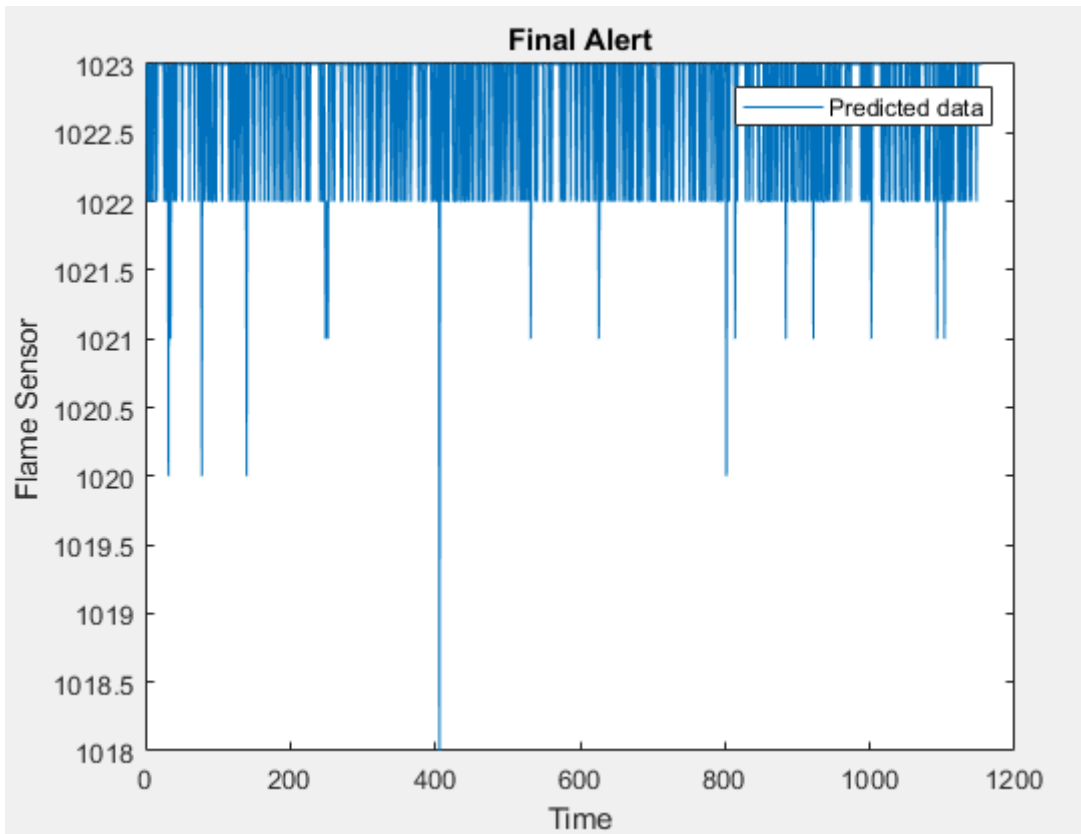
Treino sem chama e confirmação com chama (Variação 1):

-Alerta Final:



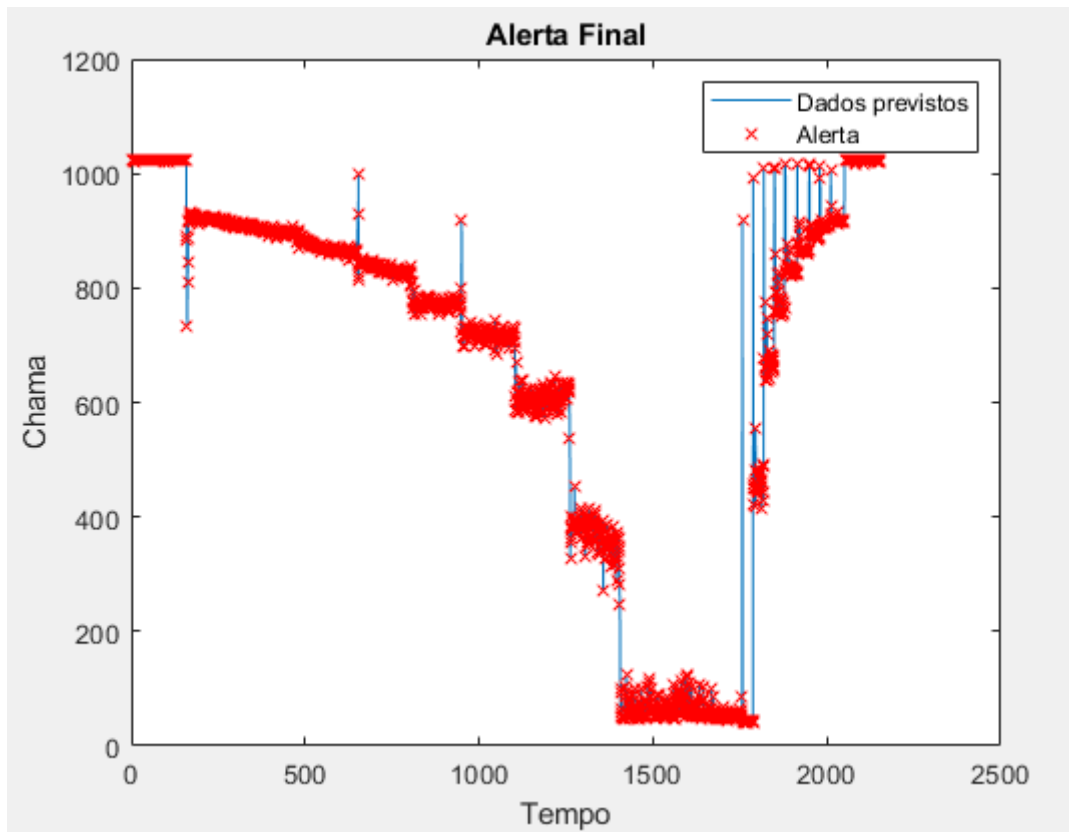
Treino com chama e confirmação sem chama (Variação 2):

-Alerta Final:



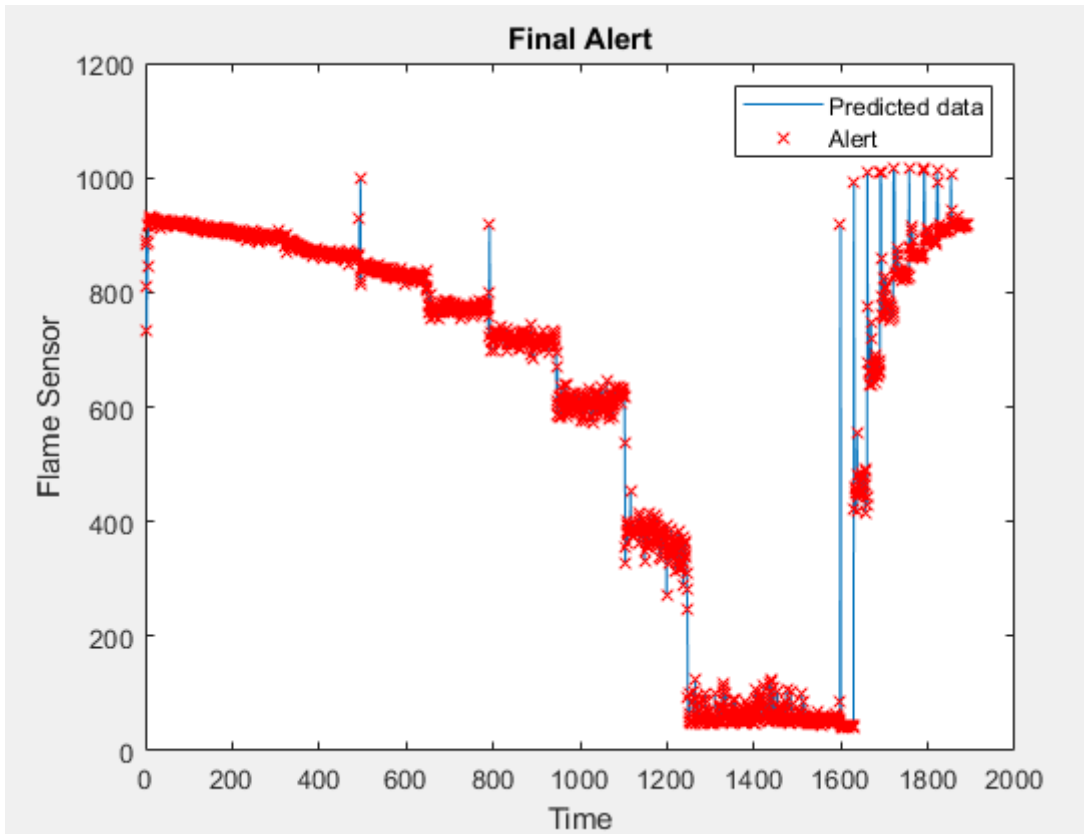
Treino com todos os valores:

-Alerta Final:



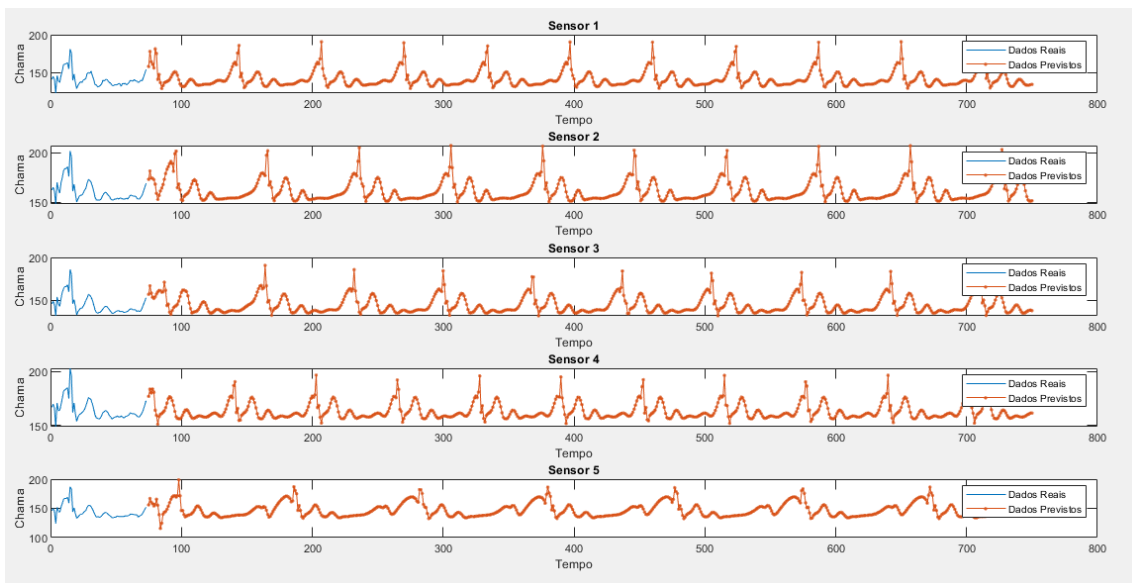
Treino com 50% sem chama e 50% com chama (Variação 4):

-Alerta Final:

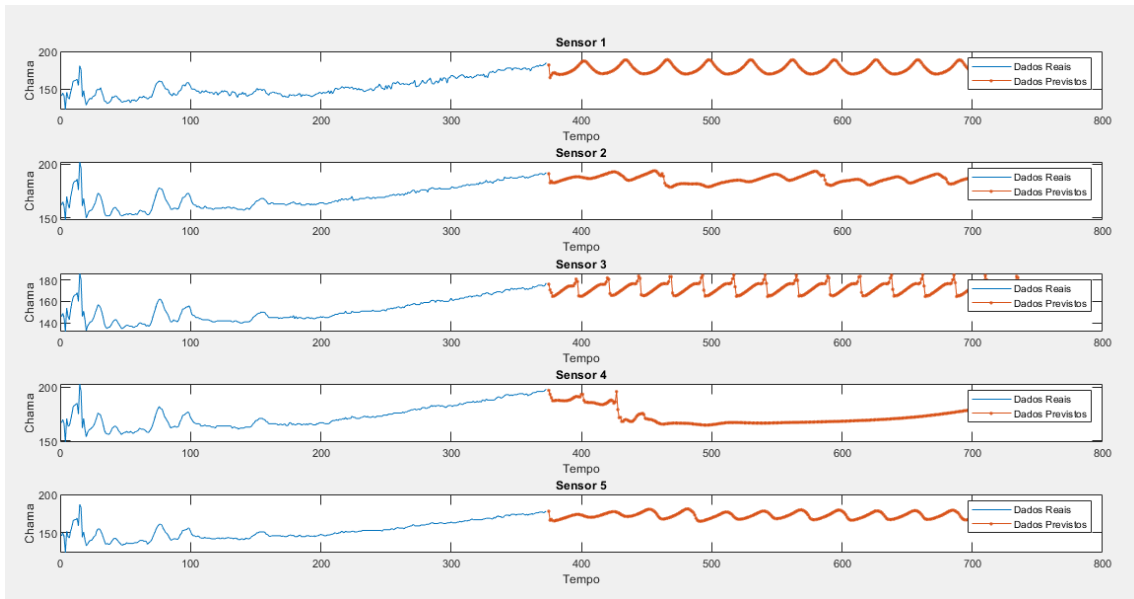


Simulação II

Previsão com 10% de dados da base de treino:

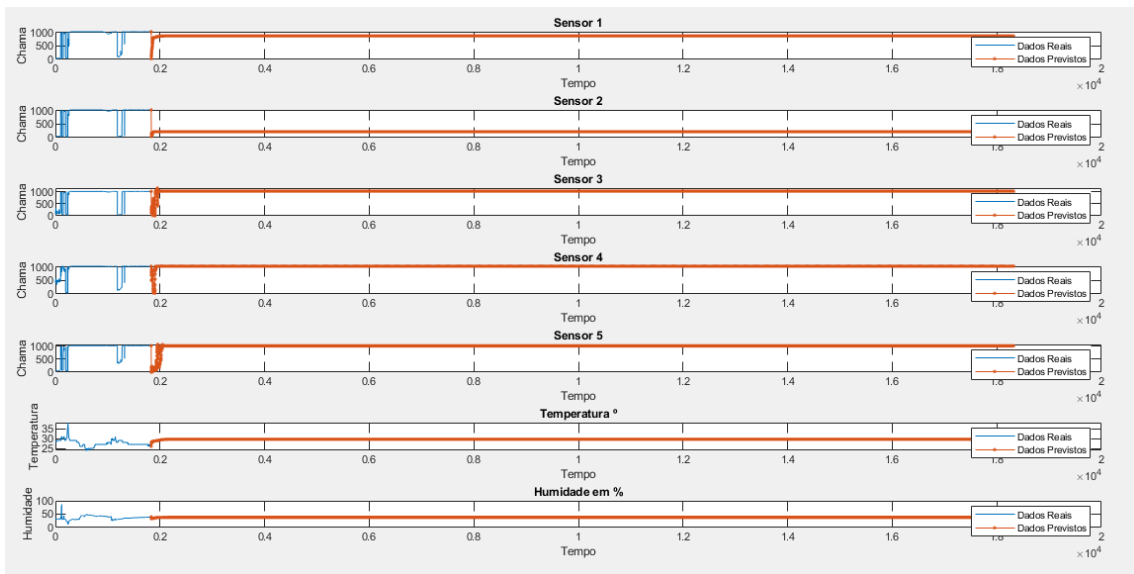


Previsão com 50% de dados da base de treino:

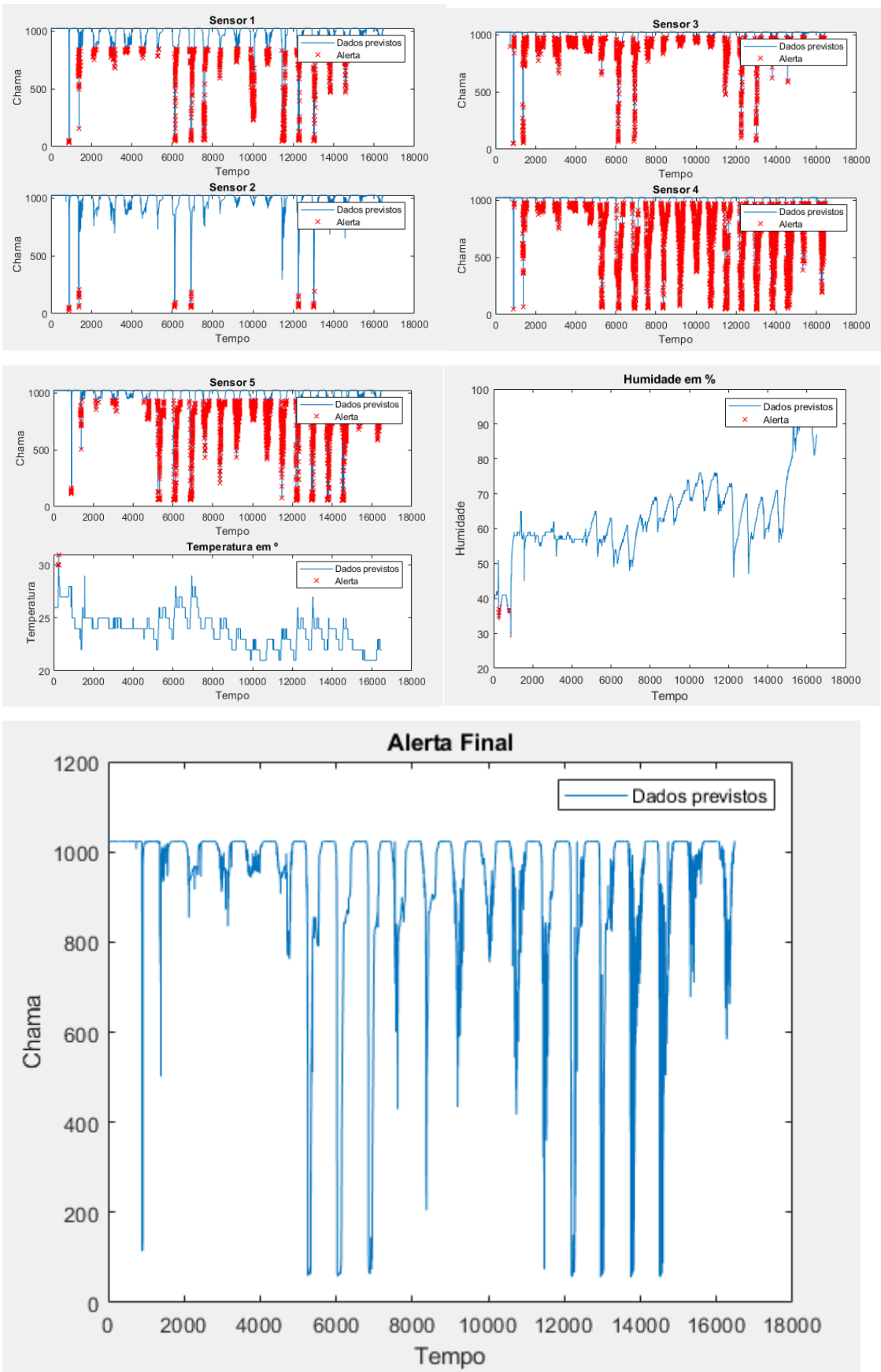


Simulação III

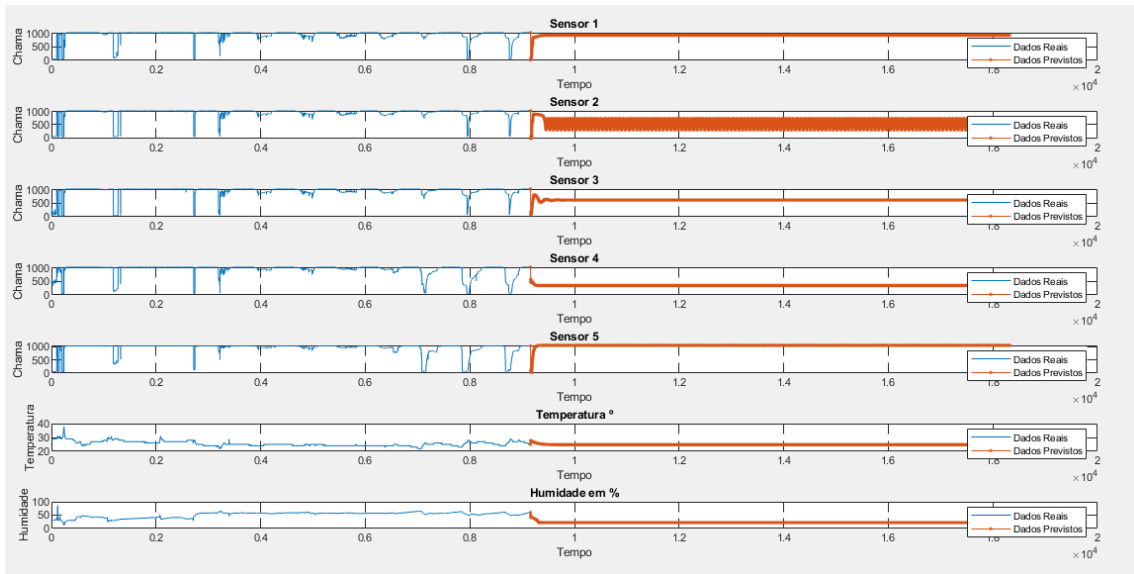
Previsão com 10% de dados da base de treino:



Alertas individuais e alerta final:



Previsão com 50% de dados da base de treino:



Alertas individuais e alerta final:

