

pmSensing: Uma Rede de Sensoriamento Participativo para Monitoramento Preditivo de Material Particulado *

Lucas L. S. Sachetti¹, Enzo B. Cussuol¹, José Marcos S. Nogueira², Vinicius F. S. Mota¹

¹Departamento de Informática – Universidade Federal do Espírito Santo
Vitória – Brasil

{lucas.sachetti, enzo.cussuol}@edu.ufes.br vinicius.mota@inf.ufes.br

²Departamento de Ciência da Computação – Universidade Federal de Minas Gerais
Belo Horizonte – Brasil

jmarcos@dcc.ufmg.br

Abstract. *This work presents a proposal for a wireless sensor network for participatory sensing, with IoT sensing devices developed especially for monitoring and predicting air quality, as alternatives of high cost meteorological stations. The system, called pmSensing, aims to measure particulate material. A validation is done by comparing the data collected by the prototype with data from stations. The comparison shows that the results are close, which can enable low-cost solutions to the problem. The system still presents a predictive analysis using recurrent neural networks, in this case the LSTM-RNN network, where the predictions presented high accuracy in relation to the real data.*

Resumo. *Este trabalho apresenta uma proposta de uma rede de sensores sem fio para sensoriamento participativo, com dispositivos IoT de sensoriamento desenvolvidos especialmente para monitoramento e previsão da qualidade do ar, como alternativa a estações meteorológicas de alto custo. O sistema, batizado de pmSensing, objetiva fazer a medição de material particulado. Uma validação é feita comparando os dados coletados pelo protótipo com dados das estações. A comparação mostra que os resultados são próximos, o que pode viabilizar soluções de baixo custo para o problema. O sistema ainda apresenta uma análise preditiva utilizando redes neurais recorrentes, no caso a rede LSTM-RNN, onde as previsões apresentaram alta acurácia em relação aos dados reais.*

1. Introdução

A poluição atmosférica tem sido uma preocupação constante para a população. Por conta disso, o monitoramento de dados sobre qualidade do ar tem aumentado gradativamente. Os principais poluentes monitorados são Monóxido de Carbono (CO), Ozônio (O₃), Dióxido de Nitrogênio (NO₂), Dióxido de Enxofre (SO₂) e o Material Particulado - MP (do inglês, *Particulate Matter - PM*) [Roy et al. 2017, dos Santos et al. 2018]. O monitoramento desses poluentes geralmente é feito em macroescala por estações meteorológicas em pontos estratégicos. No entanto, estações meteorológicas possuem alto custo de instalação e manutenção [IEMA 2021].

*O presente trabalho foi realizado com apoio financeiro da FAPES, FAPEMIG, FAPESP (Grants #2018/23011-1 e #2020/05182-3), CNPq e e Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil(CAPES) - Código 001.

Trabalhos recentes propõem o sensoriamento participativo e/ou Internet das coisas para monitoramento dos gases poluentes com um baixo custo [Liu et al. 2015, dos Santos et al. 2018]. Entretanto, a coleta de material particulado é geralmente deixada a cargo de sensores de alto custo, como os disponíveis em estações meteorológicas. O material particulado varia em quantidade, tamanho, formato e composição química, sendo mensurado como partículas com diâmetro de: até $10\ \mu\text{m}$ (PM_{10}), inferiores a $2.5\ \mu\text{m}$ ($\text{PM}_{2.5}$) e menores que $1\ \mu\text{m}$ ($\text{PM}_{1.0}$) [Abbey et al. 1999]. Quanto menores as partículas mais profundamente elas adentram no sistema respiratório, causando danos à saúde, que dependerão da composição do material particulado [WHO 2016]. Contudo, os sensores de materiais particulados capturam qualquer tipo de partícula do ar, mesmo as atóxicas como polens, por exemplo. Por este motivo, este trabalho considera que o aumento de sensores de material particulado em uma região pode minimizar o problema causado pelo sensoriamento errado de apenas um sensor. Para isto, é necessário que a rede de sensores de material particulado tenha um baixo custo mas alta confiabilidade.

Este trabalho apresenta uma proposta de uma rede de sensores sem fio para sensoriamento participativo, com dispositivos IoT de sensoriamento desenvolvidos especialmente para monitoramento e predição da qualidade do ar. No caso deste artigo, o dispositivo sensor, batizado de *pmSensing*, visa especificamente a coleta de material particulado para o monitoramento e predição de sua ocorrência no ambiente monitorado. Para tal, foi desenvolvido um protótipo do dispositivo utilizando sensores de baixo custo, capacidade processamento e armazenamento, operação isolada ou conectada à Internet com interface Wi-Fi para acesso a plataformas em nuvem. Devido ao baixo custo, pode-se aumentar o número de sensores pela cidade, permitindo até mesmo que os próprios cidadãos possam instalá-los em suas residências, criando assim, uma rede de sensoriamento e monitoramento participativos. Desta forma, o *pmSensing* visa possibilitar uma abrangência mais ampla e até mesmo móvel ou dinâmica de pontos de uma cidade, permitindo uma visão em microescala da qualidade do ar, na direção da criação de cidades inteligentes. Por exemplo, o *pmSensing* poderia ser anexados aos ônibus e/ou bicicletas compartilhadas da cidade para o monitoramento contínuo e de forma participativa da qualidade do ar. Além da rede de sensoriamento, a arquitetura proposta visa fazer o monitoramento ativo da qualidade do ar, isto é, verificar que a quantidade de material particulado em um determinado horário está dentro do esperado. Por exemplo, autoridades poderiam diminuir a circulação de veículos em horários previstos para alta poluição em uma região.

Inicialmente, foi realizada uma validação do protótipo comparando o monitoramento efetuado pela rede *pmSensing* com o monitoramento realizado por sensores em estações meteorológicas. Após isso, o protótipo do dispositivo foi disponibilizado para usuários não especialistas fazerem o monitoramento do ambiente de suas residências. Os dados monitorados eram enviados periodicamente para uma plataforma em nuvem. Em seguida, foi analisada a acurácia de modelos de predição a partir dos dados coletados pelo *pmSensing*. As principais contribuições deste trabalho são:

- Proposição e concepção de uma rede de sensores com monitoramento participativo para monitoramento ambiental de material particulado;
- Desenvolvimento de um protótipo de sensoriamento com capacidade de coleta, armazenamento, pré-processamento e comunicação sem fio.
- Demonstração de que a utilização de sensores de baixo custo e utilizados por usuários voluntários pode chegar a resultados semelhantes aos obtidos por senso-

res de alto custo de estações meteorológicas.

O restante deste trabalho está organizado como se segue: A Seção 2 apresenta uma fundamentação teórica juntamente com trabalhos relacionados. A Seção 3 apresenta o desenvolvimento, o funcionamento, a validação e a implantação do *pmSensing*. A Seção 4 apresenta uma avaliação do sistema. A Seção 5 mostra uma análise preditiva utilizando LSTM. A Seção 6 conclui o estudo discutindo resultados, limitações e trabalhos futuros.

2. Fundamentação Teórica e Trabalhos Relacionados

Esta seção apresenta uma base teórica acerca dos temas relacionados ao trabalho, além de trabalhos correlatos que foram estudados durante o desenvolvimento.

2.1. Qualidade do ar

A crescente urbanização em todo o globo terrestre tem gerado aumento no consumo de energia e na emissão de poluentes, sejam estes oriundos da queima de combustíveis fósseis, sejam oriundos da produção industrial.

Os principais poluentes monitorados que afetam a saúde das pessoas são o óxido de nitrogênio (NO₂ ou NO_x), monóxido de carbono (CO), dióxido de enxofre (SO₂), ozônio (O₃) e o Material Particulado. O interesse nos efeitos da poluição do ar relacionados à saúde tornaram-se mais expressivos depois que dois estudos de um grupo dos EUA sugeriram que a exposição à poluição do ar estava associada ao encurtamento da vida das pessoas [Dockery et al. 1993]. Em especial, já foram demonstradas relações significativas entre materiais particulados com diâmetro inferior a 10 μ m (PM₁₀) e doenças respiratórias [Pope III et al. 1991, Roemer et al. 1993] e até com mortes por doenças respiratórias não malignas em homens e mulheres, bem como com a mortalidade por câncer de pulmão em homens não fumantes [Abbey et al. 1999].

Existem políticas públicas para o monitoramento de gases e materiais particulados utilizando estações meteorológicas. No estado do Espírito Santo, o Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA) disponibiliza um índice de qualidade do ar de acordo com as concentrações dos poluentes, exibidos na Tabela 1 em μ g/m³ (micrograma por metro cúbico). A relação entre a quantidade de poluentes mensurados pelos sensores e os efeitos sobre a saúde são apresentados na Tabela 2. Vale ressaltar que a base de dados coletados pelo IEMA é pública e o direcionamento deste trabalho no monitoramento de MP_{2,5} (PM_{2,5}) deve-se também à disponibilidade desses dados.

Classificação	Concentração					
	MP ₁₀ 24h	MP _{2,5} 24h	SO ₂ 24h	NO ₂ 1h	O ₃ 8h	CO 8h
Boa	0 - 50	0 - 25	0 - 20	0 - 200	0 - 100	0 - 10.000
Moderada	>50 - 120	>25 - 60	>20 - 60	>200 - 240	>100 - 140	>10.000 - 13.000
Ruim	>120 - 150	>60 - 125	>60 - 365	>240 - 320	>140 - 160	>13.000 - 15.000
Muito Ruim	>150 - 250	>125 - 210	>365 - 800	>320 - 1130	>160 - 200	>15.000 - 17.000
Péssima	>250	>210	>800	>1.130	>200	>17000

Tabela 1. Qualidade do ar por concentração de poluentes[IEMA 2021].

Qualidade	Significado
Boa ou Moderada	Pessoas de grupos sensíveis (crianças, idosos e pessoas com doenças cardiorrespiratórias) podem apresentar sintomas como tosse seca e cansaço. A população geral não é afetada.
Ruim	Toda a população pode apresentar sintomas como tosse seca, cansaço, ardor nos olhos, nariz e garganta. Pessoas de grupos sensíveis podem ter efeitos mais sérios na saúde.
Muito Ruim	Toda a população pode apresentar agravamento dos sintomas como tosse seca, cansaço, ardor nos olhos, nariz e garganta e ainda falta de ar e respiração ofegante. Efeitos ainda mais graves à saúde de grupos sensíveis.
Péssima	Toda a população pode apresentar riscos de manifestações de doenças respiratórias e cardiovasculares. Aumento de mortes prematuras em pessoas de grupos sensíveis.

Tabela 2. Relação entre a qualidade do ar e os riscos à saúde populacional

2.2. Predição em séries temporais

Os dados coletados pelos sensores formam um conjunto de observações ordenadas no tempo, chamadas de séries temporais [Azzouni and Pujolle 2017]. Os dados coletados pelo sensoriamento proposto neste trabalho são tratados como uma série temporal, o que possibilita a utilização de métodos analíticos para calcular a qualidade do ar, se boa ou não. Entre esses métodos tem-se: análise descritiva, que apenas descreve os eventos atuais e passados; análise preditiva, que utiliza dados passados para prever ações futuras; análise prescritiva, para tomar ou recomendar decisões baseadas nos resultados obtidos pelas análises preditiva e/ou descritiva. No contexto deste trabalho, a análise preditiva possibilita o monitoramento ativo da qualidade do ar, identificando observações fora dos valores esperados, possibilitando, por exemplo, a geração de alertas para autoridades competentes.

Existem diversas formas de se fazer predição de dados, entre elas a utilização de redes neurais artificiais (ANN, do inglês *Artificial Neural Networks*) [Crone and Kourentzes 2010], comumente utilizadas para reconhecimento de padrões. Uma rede neural recorrente (RNN, do inglês *Recurrent Neural Networks*) utiliza dados que possuem uma forte relação com o tempo [Boné et al. 2003]. As RNNs utilizam informações anteriores para servir de base ao dado atual na rede neural. Quando os dados são fortemente relacionados com o passado recente, as redes recorrentes de memória longa de curto prazo (LSTM, do inglês *Long Short-term Memory Recurrent Neural Network*) são mais eficazes para realizarem predições [Azzouni and Pujolle 2017]. Uma LSTM não só retém as características de temporização da estrutura RNN, como também tem a função de memória para séries temporais.

A ideia original LSTM foi inicialmente proposta em [Hochreiter and Schmidhuber 1997] e desde então diferentes modelos têm sido propostos para melhorar seu desempenho [Cho et al. 2014], [Li et al. 2019], [Karevan and Suykens 2020], [Hu et al. 2020]. As LSTMs utilizam um mecanismo de bloqueio para definir os dados, sua duração e tempo de leitura na célula de memória.

A rede neural recorrente LSTM consiste em um módulo com células de memória que podem aprender características de dados no domínio do tempo. O módulo de memória na rede neural recorrente LSTM contém três portas (*gates*): uma porta de entrada,

uma porta de esquecimento e uma porta de saída. Essas portas controlam a entrada, a atualização e a saída das informações, respectivamente, para que a rede tenha uma determinada função de memória. A saída é um número entre 0 e 1, no qual 1 indica a manutenção de todas as informações de entrada, e 0 indica a remoção dessas informações. Este trabalho propõe o uso de LSTMs para fazer a predição de material particulado a partir de séries temporais coletadas pelos sensores do sistema proposto.

2.3. Trabalhos relacionados

São amplamente discutidas na literatura propostas de aplicações, protocolos e análises de desempenho relacionadas a Internet das Coisas aplicada ao monitoramento da qualidade do ar, principalmente em cidades inteligentes [da Lima and da Silva 2015, dos Santos et al. 2018].

[da Lima and da Silva 2015] desenvolveram um sensor de baixo custo para monitorar a qualidade do ar enviando as informações via comunicação *Bluetooth* para um celular rodando o sistema Android, para medição de concentrações de NH₃, NO_x, benzeno, fumaça e CO₂. No trabalho [Liu et al. 2015] foram utilizadas bicicletas públicas para coletar dados de poluição na cidade medindo gases e partículas, utilizaram tecnologia *Bluetooth* (FBT06M) para enviar dados coletados para uma estação base.

Em [dos Santos et al. 2018] foi utilizado um kit de sensores instalados em ônibus de uma cidade para identificar as regiões com maiores índices de poluição. Foram medidos os gases O₃, NO₂, CO₂, a temperatura ambiente e a umidade do ar. Um mapa da cidade foi construído mostrando a intensidade da poluição em cada ponto da rota do ônibus. Em [Al-Ali et al. 2010], os autores criaram um sistema com um dispositivo com sensores para medir a qualidade do ar (CO, NO₂ e SO₂) de uma cidade, enviar os dados via rede móvel para um servidor e construir um mapa da poluição. O kit utiliza um modem GPRS para enviar os dados para um servidor que utiliza uma API do Google Maps. As propostas iniciais do *pmSensing* eram justamente a instalação do protótipo em ônibus e bicicletas compartilhadas, porém devido à questões burocráticas e o curto prazo elas foram descontinuadas.

Com relação à predição de concentração de PM_{2.5} utilizando redes LSTM, que é uma técnica utilizada também neste trabalho, destacam-se os trabalhos de [Thaweephol and Wiwatwattana 2019] e [Li et al. 2020]. Em [Thaweephol and Wiwatwattana 2019], os autores treinaram um modelo de rede LSTM utilizando dados de séries temporais sobre a qualidade do ar de uma estação de polícia metropolitana na cidade de Bangkok. Seus resultados foram comparados com outro modelo utilizado chamado SARIMAX (do inglês, *Seasonal AutoRegressive Integrated Moving Average with eXogenous regressor*), que é uma versão estendida do modelo tradicional ARIMA. A comparação, feita utilizando métricas de erro, mostra que o modelo SARIMAX chegou a obter o dobro de erros em relação ao LSTM. Os autores mostraram forte correlação entre o PM_{2.5} e outros poluentes, como PM₁₀, NO e CO.

Já [Li et al. 2020] elaboraram o *Attention-based CNN-LSTM*, uma variante do LSTM com redes convolucionais (CNN-LSTM) que utiliza uma camada com mecanismos de atenção para capturar as partes mais importantes da concentração de PM_{2.5} quando se considera características de estados passados. Também são mostradas algumas correlações entre as concentrações do material particulado e a temperatura do ar, umidade

e visibilidade (em km). Os autores compararam sua proposta com o *Support Vector Regression* (SVR), *Multilayer Perceptron* (MLP), *Random Forest Regression* (RFR), RNN, LSTM, e CNN-LSTM, e mostraram que obtiveram resultados satisfatórios em relação às métricas de erro *Root Mean Square Error* (RMSE) e *Mean Absolute Error* (MAE).

3. Desenvolvimento, funcionamento, validação e implantação do pmSensing

Esta seção apresenta o módulo de sensoriamento *pmSensing*, sua arquitetura, seu funcionamento, a forma como os dados brutos coletados são transformados, bem como a implantação de unidades em uma região da cidade.

3.1. Arquitetura e componentes

O módulo de sensoriamento *pmSensing*, cuja estrutura é mostrada na Figura 1, é composto por um microcontrolador com memória, um sensor de material particulado e uma interface de comunicação sem fio. Para o desenvolvimento do módulo foram levados em consideração o custo, consumo energético, memória e desempenho. Para esta versão, foi utilizado o microcontrolador NodeMCU ESP8266 e a tecnologia de rede sem fio 802.11.

Como as estações providas pela Universidade Federal do Espírito Santo (UFES) e pelo IEMA ficam em um ambiente aberto, sujeito a chuvas, ventos e sol, entre outros fatores, o protótipo foi acondicionado em uma caixa hermética. O microcontrolador e bateria foram armazenados internamente, e o sensor de material particulado foi instalado na parte de fora da caixa. As Figuras 2 e 3 mostram o protótipo que foi desenvolvido.

Para operar em ambientes externos sem possibilidade de acesso à rede internet para o envio de dados coletados, a solução utiliza um microcontrolador com memória não volátil. Para esta versão do *pmSensing* foi escolhido um microcontrolador de baixo custo energético, com capacidade de armazenamento local e com rede 802.11 disponível. A conectividade pode ser expandida utilizando outras tecnologias de comunicação, como GSM ou mesmo LoraWan. Já o armazenamento local precisa ser suficiente para coletar dados durante o período offline. A periodicidade de leitura e armazenamento dos dados é um parâmetro configurável do sistema proposto.

3.2. Funcionamento

O sistema opera de modo offline, para o monitoramento de ambientes sem conexão à internet, e de modo online, para ambientes protegidos do tempo e que possuam redes sem fio. O modo offline é também utilizado para validar os sensores com as estações que estão situadas em ambientes sem conexão. A validação permite verificar se o monitoramento do *pmSensing* se aproxima dos resultados das estações.

No modo online, para operação em ambientes protegidos, os dados são enviados via rede para um servidor web, utilizando acesso sem fio existente no local. Os dados são enviados para uma plataforma aberta que, neste trabalho, é a plataforma ThingSpeak [MathWorks 2017] para armazenamento e análise dos dados.

Estando o módulo de sensoriamento no modo online, os dados coletados e pré-processados no próprio módulo são enviados para a plataforma ThingSpeak para serem armazenados e processados. Essa plataforma permite a criação de canais que recebem dados via web socket ou via o protocolo MQTT. A cada 30 segundos os dados de uma leitura do sensor de material particulado são enviados para a plataforma. Esse tempo é o intervalo mínimo para envio de mensagens definido pela plataforma no plano gratuito.

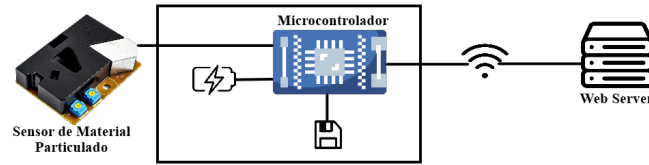


Figura 1. Arquitetura do módulo de sensoriamento

3.3. Coleta e pré-processamento dos dados coletados

O sensor de material particulado utilizado foi o sensor de poeira Samyoung DSM501, que é um sensor de densidade de partículas de baixo custo e de tamanho compacto, fabricado pela empresa coreana Samyoung Electronics. O sensor possui um aquecedor (resistor) para gerar o calor que cria uma corrente de ar ascendente que atrai o ar externo para o interior do módulo. O DSM501 mede o intervalo de tempo em que uma partícula é detectada pelo sensor. Esse tempo é chamado de *Low Pulse Occupancy*, que pode ser visto também como a porcentagem de opacidade do ar que circula pelo sensor. Desse modo, ele detecta poeira, partículas e pólen no ar ($PM_{2.5}$ e $PM_{1.0}$) em ambientes fechados. O *pmSensing* monitora o $PM_{2.5}$, sendo assim possível validar e comparar os dados obtidos com os dados das estações do IEMA.

Para que a comparação dos dados possa ocorrer é preciso converter os dados sensorizados para outro formato, de concentração em microgramas por metro cúbico, no caso denominado *Low Ratio*. De acordo com o *datasheet* do fabricante, o dado bruto obtido pelo sensor é na forma de *Low Pulse Occupation*. A conversão é feita no próprio módulo. A Equação 1, converte o *Low Pulse Occupation* para uma concentração de partículas por litro (pcs/L) ou partículas por 0.01 pés cúbicos (pcs/0.01 cf) (*Low ratio [%]*).

$$Concentration = 1.1 \times (Low\ ratio^3) - 3.8 \times (Low\ ratio^2) + 520 \times Low\ ratio + 0.62 \quad (1)$$

Essa concentração é possível de ser convertida em concentração por metro cúbico ($\mu g/m^3$), aplicando a Equação 2, conforme [Roy et al. 2017].

$$Particle_{Mass} = Density \times Volume \quad (2)$$

$$PM_{Concentration}(\mu g/m^3) = Number_Of_Particle \times 3531.5 \times Particle_{Mass}$$

Importante ressaltar que [Roy et al. 2017] assumem que todas as partículas dessa concentração são esféricas e possuem densidade igual a 1.65×10^{12} e que o raio de cada partícula, no caso $PM_{2.5}$, é igual a $0.44 \mu m$. A massa é composta pela multiplicação da densidade pelo volume, e a concentração é obtida a partir da multiplicação dessa massa pela quantidade de partículas e por um fator de conversão de medida.

3.4. Implantação das unidades de sensoriamento

Os sensores foram implantados em locais próximos as estações do Instituto Estadual de Meio Ambiente e Recursos Hídricos (IEMA) do Espírito Santo na Grande Vitória-ES.



Figura 2. Protótipo do sensor fechado

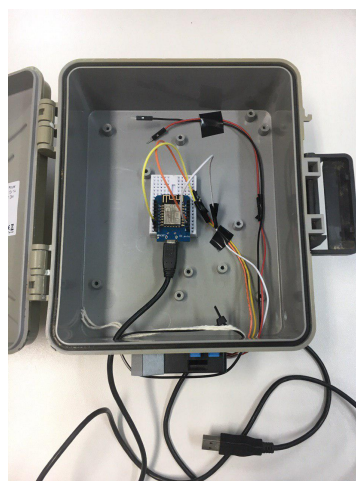


Figura 3. Protótipo do sensor aberto

Desta forma, os dados coletados foram comparados com os dados coletados pelos sensores em macroescala do IEMA. Os dados disponibilizados pelo IEMA são de coletas feitas a cada 60 minutos; dessa forma, para permitir uma validação ajustada, foi calculada a média da concentração de material particulado para cada 60 minutos.

Para avaliar o sensoriamento participativo, foram desenvolvidas quatro unidades do protótipo e distribuídas a voluntários. Para coletar dados em mais locais, os protótipos eram entregues a novos voluntários. A Figura 4 apresenta os locais nos quais os sensores foram instalados para a coleta de dados. Quanto mais pessoas utilizando numa região, melhor será a mediana dos resultados que serão mais fidedignos aos resultados das estações meteorológicas na qual são usados para validação.

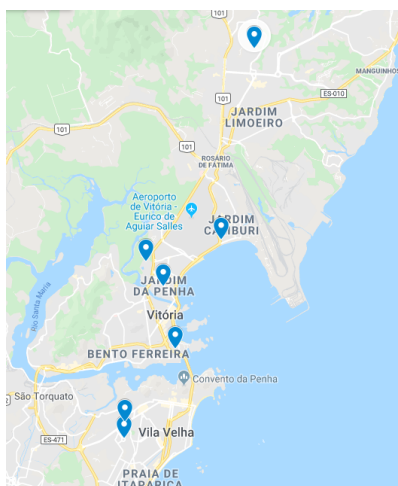


Figura 4. Mapa dos locais de monitoramento

4. Avaliação do pmSensing

O protótipo do dispositivo, construído com sensores de materiais particulados de baixo custo, foi avaliado pela comparação com medições de materiais particulados feitas por

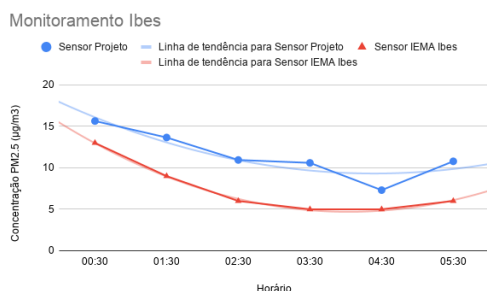


Figura 5. Monitoramentos de $PM_{2.5}$ no bairro Ibes

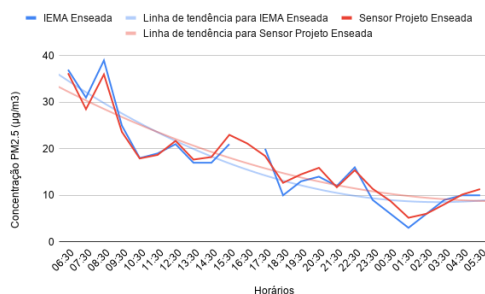


Figura 6. Monitoramentos de $PM_{2.5}$ na Enseada do Suá

estações meteorológicas. Os dados coletados pelas unidades de sensoriamento em diversos pontos da Grande Vitória, com auxílio de voluntários, foram avaliados e classificados de acordo com o Índice de Qualidade do Ar (IQA), disponibilizado pelo IEMA. Os sensores foram colocados em casas próximas às estações de monitoramento do IEMA.

4.1. Validação

Em testes preliminares, observou-se que o posicionamento do protótipo tem relevância nos resultados. Contudo, desde que colocado em posições expostas ao tempo, como próximo a uma janela ou em uma varanda, os resultados são satisfatórios em relação aos dados observados pelas estações meteorológicas, conforme apresentado a seguir.

O primeiro teste de validação ocorreu no mês de maio de 2019 (28/05/2019) na localidade próxima ao sensor de macroescala do IEMA no bairro Ibes em Vila Velha. A Figura 5 apresenta os dados de concentração de $PM_{2.5}$ em $\mu g/m^3$ coletados pelo *pmSensing* em comparação com os dados do sensor macroescala do IEMA. Observa-se que o *pmSensing* obteve uma concentração de $3.22 \mu g/m^3$ em média em comparação à concentração mensurada pelo IEMA. Embora haja essa diferença da leitura, ambos os dados monitorados estão na classificação BOA no Índice de Qualidade do Ar, de acordo com a tabela 1. Ademais, observa-se que a curva de tendência tem comportamento semelhante.

O segundo teste de validação foi feito nas proximidades da estação do IEMA na Enseada do Suá, em Vitória. Este teve uma duração de 24h e ocorreu nos dias 28 e 29 de agosto de 2019. O *pmSensing* foi instalado em uma varanda de um apartamento do sexto andar de um prédio ao lado do Corpo de Bombeiros onde o sensor do IEMA estava instalado. A Figura 6 apresenta o resultado do monitoramento realizado pelo *pmSensing* e do monitoramento do IEMA referente àquele dia. Observa-se que os dados coincidem em muitos dos pontos observados e a diferença média entre as séries foi de 0.53. A base de dados para o dia da coleta estava incompleta, e por isso houve uma descontinuidade da curva dos dados do IEMA neste ponto (16h30). Um dos desafios da comparação com a base do IEMA é que os dados de um determinado mês são liberados apenas no mês seguinte. Portanto, o *pmSensing* realizou coleta de dados em dias específicos mas pode comparar os dados apenas no mês seguinte à coleta.

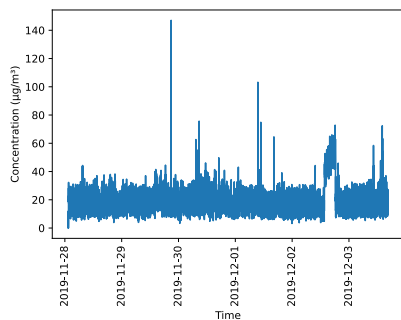


Figura 7. Concentração de PM (Jardim Camburi)

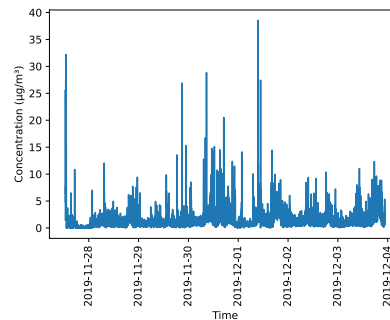


Figura 8. Concentração de PM (Jardim da Penha)

4.2. *pmSensing* como um sistema de Sensoriamento participativo

Após validar a eficácia do *pmSensing* comparando-o com os sensores de macroescala do IEMA, os protótipos foram utilizados para monitorar o índice de qualidade do ar em ambientes domésticos. Este experimento conduzido na região da Grande Vitória teve grande aceitação de voluntários devido a um questionamento popular sobre a incidência de poeira e pó de minério próximos à área portuária de escoamento de minério na região. Nos bairros próximos a esta área portuária, o IEMA faz o monitoramento apenas de PM_{10} e outros gases. Porém, o protótipo do *pmSensing* monitora apenas $PM_{2.5}$ e $PM_{1.0}$. Portanto, não houve comparação com dados do IEMA, mas sim com dados coletados concomitantemente em diferentes bairros.

Uma comparação a ser destacada é entre os bairros Jardim Camburi e Jardim da Penha. A Figura 7 mostra os níveis de material particulado entre 28 de novembro e 03 de dezembro. A média diária nesse período foi de $28.745 \mu g/m^3$, o que é considerado MODERADO no IQA no bairro Jardim Camburi. Também é possível ver que houve um pico de algumas horas acima de $40 \mu g/m^3$, o que é ainda considerado MODERADO e poucos picos acima de $60 \mu g/m^3$, considerados RUIM no IQA. Vale ressaltar que Jardim Camburi é o bairro mais próximo ao porto exportador de minério de ferro. Por outro lado, o bairro Jardim da Penha, afastado ~ 7 km da área portuária, foram feitas medições ao mesmo tempo que no Jardim Camburi. A Figura 8 mostra o monitoramento no mesmo espaço de tempo do anterior, quando a média diária de concentração foi de $3.920 \mu g/m^3$, sendo possível notar também alguns picos acima de $25 \mu g/m^3$. Apesar disso, o IQA neste local foi considerado BOA.

Com finalidade de verificar a sazonalidade dos dados, foram destacados alguns dias de monitoramento de um mesmo local. A Figura 9 mostra o monitoramento no Jardim da Penha no dia 28 de novembro, e é possível verificar picos em torno das 9h e uma crescente de picos após as 18h. Essa mesma crescente de picos também é observada em outros dias, como o da Figura 10 que se refere ao dia 03 de dezembro no mesmo local.

A Tabela 3 apresenta o índice de qualidade do ar monitorado em cada local na data do experimento. Vale ressaltar que o monitoramento foi feito em ambientes domésticos e não deve ser generalizado para a região inteira, a priori.

Local	PM2.5 ($\mu\text{g}/\text{m}^3$)	IQA
Aribiri	3.224	BOA
Colina de Laranjeiras	1.042	BOA
Enseada do Suá	17.178	BOA
Ibes	11.369	BOA
Jardim Camburi	28.745	MODERADA
Jardim da Penha	3.920	BOA
UFES	1.515	BOA

Tabela 3. Índices de qualidade do ar monitorado em diversos locais

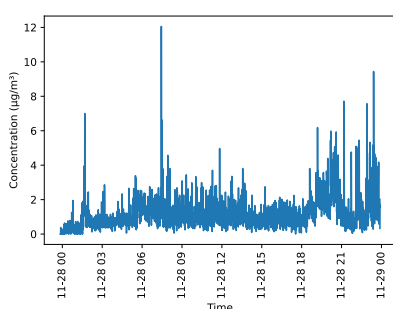


Figura 9. Concentração de PM dia A (Jardim da Penha)

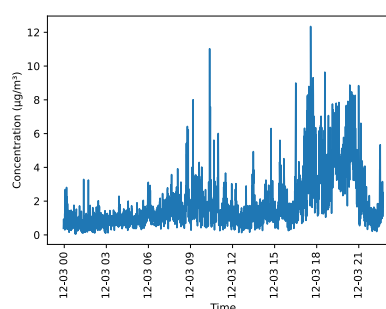


Figura 10. Concentração de PM dia B (Jardim da Penha)

5. Análise Preditiva

Com o objetivo de realizar o monitoramento preditivo, foi analisado se modelos preditivos se adéquam aos dados de material particulado. O monitoramento preditivo visa prever se a quantidade de material particulado em um determinado horário está dentro do esperado, e caso ocorra anomalias o sistema poderia emitir alertas, por exemplo. Tal monitoramento, pode auxiliar nas tomadas de decisões de políticas ambientais. Por exemplo, saber previamente os níveis de poluição em um local pode ajudar na tomada de decisão de fluxo de veículos em Cidades Inteligentes. Nesta seção é apresentado um modelo de previsão que utiliza os dados coletados pelo *pmSensing*.

5.1. Definição do problema

A previsão de $\text{PM}_{2.5}$ é um problema de regressão linear e pode ser resolvido com algoritmos de aprendizado de máquina supervisionados. Para prever a concentração de poluentes para a próxima hora, $t + 1$, são tomados não apenas o valor atual da concentração, t , como também seus valores anteriores, $t - 1$, $t - 2$, ..., $t - N$, onde N é o tamanho da memória, que neste trabalho foi para conter 9 leituras de concentração. Portanto são lidas 9 concentrações anteriores para tentar prever a próxima concentração. Para fazer uma previsão de material particulado a cada 5 minutos, a previsão é feita com base nos dados coletados nos 4 minutos e meio anteriores.

5.2. Arquitetura e hiperparâmetros

Foi desenvolvido então uma rede neural utilizando o modelo LSTM. O modelo utilizado possui apenas uma *hidden layer* de LSTM, pois como o problema tratado possui apenas uma variável foi decidido por utilizar apenas uma camada por ser um problema mais simples de tratar. Aumentar o número de camadas em problemas assim pode não levar a ganhos tão consideráveis. O número de nós/neurônios internos variou entre 25 e 100.

Entre as funções de ativação, foram utilizadas ReLu, Sigmoid, Tanh e Softmax para verificar a eficácia de cada uma. Além disso foi iterado também o número de *epochs*, sendo utilizado 100, 200, 500 e 1000. Foi utilizado apenas o *optimizer* Adam.

Para executar os algoritmos, usou-se um computador com uma CPU Intel Core i5-4210U 1.70GHz, com gráficos integrados, 4GB de memória RAM e 200 GB de armazenamento com sistema operacional Ubuntu 16.04.5 LTS. Foram utilizadas as bibliotecas de Tensorflow e Keras [Abadi et al. 2016] para implementação das redes neurais.

5.3. Resultados e análise

Foram utilizadas métricas de erro, como *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), e *Root Mean Squared Error* (RMSE) para analisar a acurácia do modelo. Os dados foram divididos em 70% para treino e 30% para teste. As redes foram treinadas e testadas de acordo com os hiperparâmetros utilizados Seção 5.2.

Dentre essas, a que deu melhor resultado foi a que utilizou os seguintes hiperparâmetros: 61 neurônios internos, função de ativação Softmax, 100 épocas. Os resultados das métricas foram de: 0.5166 de MAE, 1.0234 de MSE e 1.0116 de RMSE. O MAE avalia os erros relativos às previsões, já o MSE é uma medida da qualidade do estimador, onde valores próximos de 0 são melhores. O RMSE nos dá o desvio padrão do erro residual, e quanto menor o resultado melhor será a performance do modelo. A Figura 11 apresenta uma comparação do melhor modelo com os dados reais.

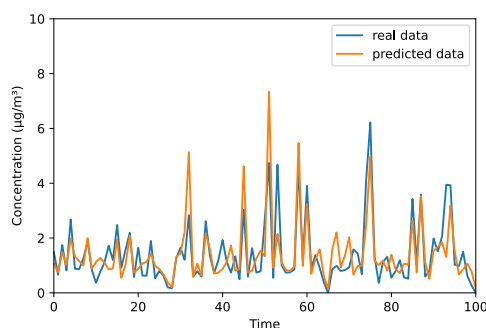


Figura 11. Comparação da previsão do melhor modelo com dados reais

6. Conclusão

Este trabalho apresentou a proposta de um sistema (*pmSensing*) para monitoramento de material particulado em microescala, pelo qual é possível avaliar a qualidade do ar de um ambiente e fazer um monitoramento preditivo dessa qualidade do ar utilizando redes neurais. O sistema pode operar tanto em modo offline quanto online. O *pmSensing* foi validado por comparação com dados coletados por estações meteorológicas. Para isso, unidades do protótipo foram instaladas próximas a estações meteorológicas do IEMA que monitoram material particulado (PM_{2.5}). Os resultados mostraram que dispositivos sensores de baixo custo (cerca de R\$120,00), como é o caso deste trabalho, podem obter resultados semelhantes aos de sensores de estações meteorológicas de alto custo.

Unidades do protótipo foram instaladas em alguns pontos da região metropolitana de Vitória ES para monitorar a qualidade do ar, quando foi possível observar que existem

diferenças nos índices de qualidade do ar (IQA) entre bairros, muitas vezes próximos. Uma das limitações do protótipo é em relação ao ambiente em que está sendo instalado, o que pode aumentar a quantidade de material particulado, um trabalho futuro é aumentar o número de unidades instaladas em um mesmo bairro para observar a dinâmica do IQA e minimizar eventuais leituras discrepantes em uma mesma região.

O uso de microcontroladores de baixo custo traz limitações, principalmente no poder computacional e memória. O uso do microcontrolador ESP8266 atende aos requisitos de baixo custo e comunicação Wi-Fi. Contudo, requer um roteador Wi-Fi para acesso à Internet. Nesse sentido, entre as possibilidades de evolução do dispositivo, está a adição de outras interfaces de comunicação, como LORA e Bluetooth. No primeiro caso, havendo cobertura de uma rede LoraWan, evita-se conexão via roteador Wi-Fi. No segundo caso, o dispositivo poderia transmitir os dados coletados para um aparelho celular que, por sua vez, os repassaria para um servidor. Nos dois casos, agrega-se capacidade de monitoramento participativo móvel, uma propriedade desejável do sistema.

Como continuidade do trabalho, pretende-se adicionar ao módulo de sensoria-mento o monitoramento de outras métricas do IQA como temperatura, umidade, pressão atmosférica e gases poluentes. Uma meta ambiciosa que está no horizonte da pesquisa, é a de dotar o dispositivo de capacidade de sensoriar partículas líquidas que contenham vírus, como o Sars-Cov-2, para ser usado em ambientes com possibilidade de ocorrência do vírus, como elevadores, hospitais e ônibus. Além disso, o desenvolvimento de uma *dash-board* para visualização e disparo de alertas às autoridades também é um trabalho futuro. Apresentar a dinâmica do IQA em microescala de uma mesma região pode ser benéfico tanto para políticas públicas quanto para oferta de novos serviços, como apresentação do IQA das regiões da cidade para os interessados em aquisição de imóveis.

Referências

- Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al. (2016). Tensorflow: A system for large-scale machine learning. In *12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16)*, pages 265–283.
- Abbey, D. E., Nishino, N., McDonnell, W. F., Burchette, R. J., Knutsen, S. F., Lawrence Beeson, W., and Yang, J. X. (1999). Long-term inhalable particles and other air pollutants related to mortality in nonsmokers. *American journal of respiratory and critical care medicine*, 159(2):373–382.
- Al-Ali, A., Zualkernan, I., and Aloul, F. (2010). A mobile gprs-sensors array for air pollution monitoring. *IEEE Sensors Journal*, 10(10):1666–1671.
- Azzouni, A. and Pujolle, G. (2017). A long short-term memory recurrent neural network framework for network traffic matrix prediction. *arXiv preprint arXiv:1705.05690*.
- Boné, R., Assaad, M., and Crucianu, M. (2003). Boosting recurrent neural networks for time series prediction. In *Artificial neural nets and genetic algorithms*, pages 18–22. Springer.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., and Bengio, Y. (2014). Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

- Crone, S. F. and Kourentzes, N. (2010). Feature selection for time series prediction—a combined filter and wrapper approach for neural networks. *Neurocomputing*, 73(10-12):1923–1936.
- da Lima, A. L. and da Silva, V. L. (2015). Micro sensor para monitoramento da qualidade do ar. In *Gestão, Tecnologia Industrial e Modelagem Computacional*, volume 1.
- Dockery, D. W., Pope, C. A., Xu, X., Spengler, J. D., Ware, J. H., Fay, M. E., Ferris Jr, B. G., and Speizer, F. E. (1993). An association between air pollution and mortality in six us cities. *New England journal of medicine*, 329(24):1753–1759.
- dos Santos, A. S., Faccini, L. G., Marte, C. L., and Cintra, J. P. (2018). Emprego de componentes móveis de iot para análises espaçotemporais de alta resolução para avaliar a qualidade do ar em cidades médias brasileiras. In *Anais do II Workshop de Computação Urbana (COURB 2018)*, volume 2. SBC.
- Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, 9(8):1735–1780.
- Hu, J., Wang, X., Zhang, Y., Zhang, D., Zhang, M., and Xue, J. (2020). Time series prediction method based on variant lstm recurrent neural network. *Neural Processing Letters*, pages 1–16.
- IEMA (2021). Qualidade do ar. <https://iema.es.gov.br/qualidadedoar>.
- Karevan, Z. and Suykens, J. A. (2020). Transductive lstm for time-series prediction: An application to weather forecasting. *Neural Networks*, 125:1–9.
- Li, S., Xie, G., Ren, J., Guo, L., Yang, Y., and Xu, X. (2020). Urban pm2.5 concentration prediction via attention-based cnn–lstm. *Applied Sciences*, 10(6):1953.
- Li, Y., Zhu, Z., Kong, D., Han, H., and Zhao, Y. (2019). Ea-lstm: Evolutionary attention-based lstm for time series prediction. *Knowledge-Based Systems*, 181:104785.
- Liu, X., Li, B., Jiang, A., Qi, S., Xiang, C., and Xu, N. (2015). A bicycle-borne sensor for monitoring air pollution near roadways. In *2015 IEEE International Conference on Consumer Electronics-Taiwan*, pages 166–167. IEEE.
- MathWorks (2017). Thingspeak communication library for arduino, esp8266 and esp32.
- Pope III, C. A., Dockery, D. W., Spengler, J. D., and Raizenne, M. E. (1991). Respiratory health and pm10 pollution: a daily time series analysis. *American Review of Respiratory Disease*, 144(3_pt_1):668–674.
- Roemer, W., Hoek, G., Brunekreef, B., et al. (1993). Effect of ambient winter air pollution on respiratory health of children with chronic respiratory symptoms. *American Journal of Respiratory and Critical Care Medicine*, 147(1):118–124.
- Roy, S., Bose, R., and Sarddar, D. (2017). Smart and healthy city protecting from carcinogenic pollutants. *Int. J. Appl. Environ. Sci*, 12:1661–1692.
- Thaweephol, K. and Wiwatwattana, N. (2019). Long short-term memory deep neural network model for pm2.5 forecasting in the bangkok urban area. In *2019 17th International Conference on ICT and Knowledge Engineering*, pages 1–6. IEEE.
- WHO, W. H. O. (2016). Ambient air pollution: A global assessment of exposure and burden of disease.