A COMPARATIVE STUDY ON MACHINE LEARNING APPROACHES FOR PREDICTING WORKPLACE HAZARDS IN THE PUBLIC SECTOR

K V Chandrasekhar Reddy
Research Scholar
Shri JJT University
Rajasthan.

Dr. Prasadu Peddi
Professor
Shri JJT University
Rajasthan.

Dr. Jangala Sasi Kiran
Professor
Dept Of CSE
Lords Institute of Engineering & Technology

ABSTRACT

Managing public spaces for the benefit of all citizens in reducing potential risks and ensuring the effective operation of networks, the board plays an essential role. Typical limitations of traditional methods of risk assessment include human biases and data inertia. Despite their value, historical facts and expert opinion could not paint the whole picture. Applying AI provides a state-of-the-art way to analyse large and complex data sets, which improves dynamic risk prediction. Foreseeing risks in the fields of public health and security management using AI models is central to our effort. Approaches directed and unaided learning to see how well they can identify potential dangers, such as industrial accidents, public health emergencies, and ecological disasters. Therefore, it highlighted the potential of ensemble learning as a more accurate prediction model in the field of occupational injury. In constructing the model, this study also proposed the feature optimization technique that revealed the three most important features; 'nature of injury', 'type of event', and 'affected body part' in developing model. Algorithms employ a wide variety of data sources, including as segment data, environmental sensors, and event reports, to identify trends, allowing them to achieve better anticipated accuracy than traditional risk assessment methods.

Keywords: *Machine learning (ML), public health, Public Sector, AI models, data sources, risk assessment, state-of-the-art,*

INTRODUCTION

Employees play a crucial role within each company, whether small businesses or large IT companies. The process of recruiting a candidate is often composed of a series of steps designed to optimize the candidate's choice. However, the steps used during the interview process do not guarantee the success of the choice made. Interviews can be affected by candidates stress levels, misunderstandings during the review process, and other unknown human-related variables. Machine learning (ML) algorithms are now widely used to predict possible health and safety violations. Various machine learning approaches, such as decision trees, fuzzy multiple regression, fuzzy neural networks, fuzzy Bayesian networks, fuzzy logic, and bow-tie representation, are utilised. When it comes to making precise predictions, traditional machine learning and statistical methods have their limitations. The primary cause of this issue stems from the arduous and time-consuming task of gathering character properties from extensive databases. The field of machine learning is quickly embracing deep learning techniques to automate the extraction of features from large datasets collected from different sources. Based on the data, it is evident that the dynamics are non-linear because of the complex interactions between the dependent and independent factors.

LITERATURE REVIEW

Giovanna Maria Dimitri (2023) Employee attrition is a major problem that causes many companies to incur in significant costs to find and hire new personnel. The use of machine

learning and artificial intelligence methods to predict the likelihood of resignation of an employee, and the quitting causes, can provide HR departments with a valuable decision support system and, as a result, prevent a large waste of time and resources. In this paper, we propose a preliminary exploratory analysis of the application of machine learning methodologies for employee attrition prediction. We compared several classification models with the goal of finding the one that not only performs best, but is also well interpretable, in order to provide companies with the possibility of improving those aspects that have been shown to produce the quitting of their employees. Among the proposed methods, Logistic Regression performs the best, with an accuracy of 88% and an AUC-ROC of 85%.

Subhash Monda (2023) The Internet of Things (IoT) is very versatile, as shown by its vast variety of applications, one of which is healthcare. Health data may be continuously monitored and analysed in real-time via the Internet of Medical Things (IoMT) thanks to the integration of complicated technologies. With this helpful tool, you may quickly assess the chances of pregnancy issues and take action if necessary. Remote and non-invasive monitoring of vital signs including temperature, foetal movement, blood pressure (systolic and diastolic), and heart rate is made possible with the integration of IoT devices in the system.

Mohamed Zul Fadhli Khairuddin (2022) No industry can afford to delay in determining the full scope of work-related injuries. Taking into account the enormous potential of machine learning to improve predictive analysis, this study aims to develop a model that is optimised for features and can successfully forecast the severity of injuries sustained on the job. Using an open-source database of 66,405 OSHA-supplied reports of occupational injuries, this research compares five different machine learning algorithms. Some of the models include deep learning methods such as Support Vector Machine, K-Nearest Neighbours, Decision Tree, and Random Forest. With respect to accuracy and F1-score, Random Forest outperformed all of the other models that were considered. The results demonstrated that ensemble learning may improve injury prediction in the workplace.

Juan Lu (2021) Using population-level, linked administrative health data, we sought to assess the usefulness of machine learning algorithms in alerting older patients about the risk of acute coronary syndrome and mortality one year following NSAID use. Hospitalisations or deaths due to acute coronary syndrome within one year of the first supply date were used to determine outcomes. To create prediction models for the identification of ACS and mortality, complex categorisation methodologies based on machine learning were used. The model's performance was evaluated using sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC-ROC).

Li-Ya Wu (2021) To improve Taiwan's border inspection processes for imported food commodities, we have created risk prediction models using an ensemble learning technique. Our state-of-the-art suite of prediction algorithms has greatly improved food product border control. In the long run, this helps keep the public healthy by accurately detecting non-compliant materials. In order to assess the security of each incoming shipment of food, our models—created by authorities in the subject—offer suggestions. A computer science specialist used a confusion matrix to calculate many performance measurements, such as the area under the curve, recall, positive prediction value (PPV), and the F1 score, which is the harmonic mean of PPV and recall.

Safety management

Managers in the construction business and on projects greatly value safety management systems. Executing critical safety procedures well is a must for successful construction project management. With these procedures in place, the likelihood of accidents and injuries will go down, which will save up a lot of time. More than that, you can lessen the blow of any monetary setbacks. Using safety management systems is commonplace in the construction sector due to the critical nature of maintaining a risk-free workplace. Supervisors and managers-to-be may need broader competence in addition to subject-matter knowledge. To reduce the number of accidents and deaths that occur on the job, the construction industry has put in place a very effective safety management system. One of the main goals of many methods is to increase output. In order to ensure safety, it is essential to thoroughly investigate all potential factors.

Machine learning in construction industry

When compared to other sectors, the construction industry in developing nations like India has been sluggish to embrace new technologies. Because of this stagnation, workers are facing more and more difficulties in their job. Due to the high cost and specialist knowledge generally required, acquiring the necessary resources to implement new technology may be a daunting task. The use of machine learning has the ability to greatly enhance the company's future growth, which in turn will undoubtedly benefit the company's workers, subcontractors, and customers. The ability of artificial intelligence is built upon machine learning, which is a core component of AI. This technology is really useful and has a lot of promise, therefore it's everywhere. Amazingly, these robots can learn new things and make their own forecasts while humans aren't there. Advanced software algorithms provide machine learning systems the remarkable capacity to swiftly assess data and generate accurate predictions.

Artificial intelligence and machine learning models for risk prediction

Americans, whether they work for the government or for a private company, face the daunting challenge of ever-increasing healthcare costs. Finding the people with the greatest healthcare needs and looking into methods to create a more balanced connection between treatment, cost, significance, and quality is an effective technique to enhance quality, produce better outcomes, decrease expenses, and raise life satisfaction. Those who could benefit from medical intervention are identified using probabilistic risk prediction methods. Particularly in cases when the effects on patients' lives and the associated costs are substantial, like cardiovascular disease, this becomes clear. The core idea of a clinical methodology is to assess a patient's susceptibility and then create a tailored treatment plan to alleviate, cure, or manage their disease. A critical clinical strategy for attaining equality and effectiveness in financial investment is to implement a systematic approach to patient identification, planning, and treatment.

Using AI and machine learning to reduce risk

Medical records, and test results could substantially enhance the accuracy of illness diagnoses. Genetic information, medical history, and treatment outcomes are all pieces of patient data that doctors use to choose the most effective medications and refine treatment plans. This offers further support for the idea of precision medicine, which aims to enhance treatment methods while reducing the likelihood of negative health effects. The intriguing area of computer vision uses picture analysis to help computers make better judgements,

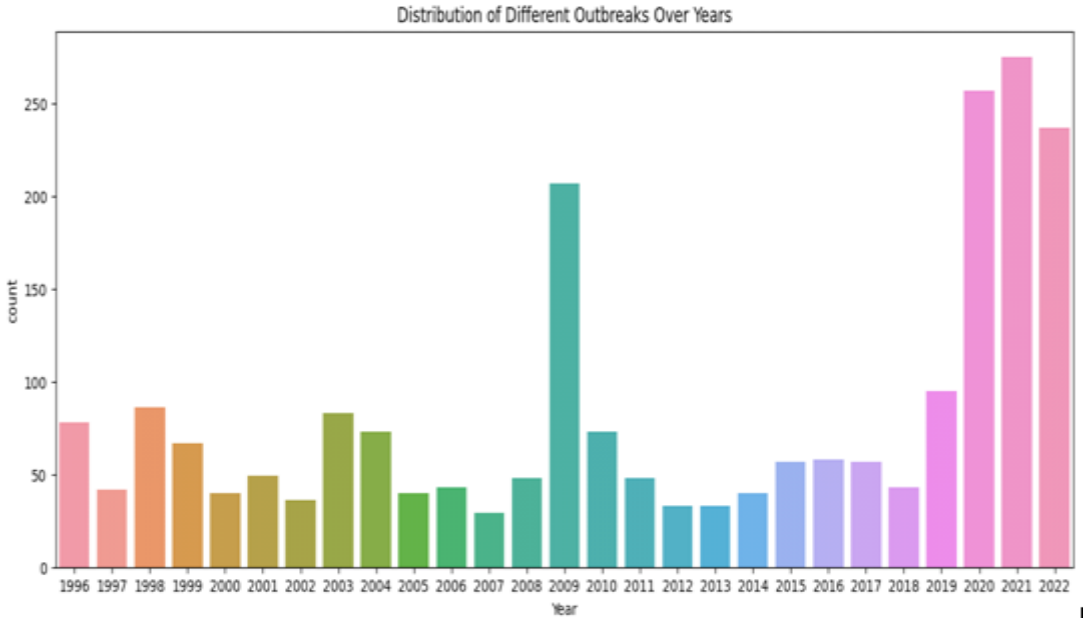
recognize patterns, and extract useful information from images. Under its wing are all of these responsibilities. Some of these applications may use DL/ML to train models on massive annotated datasets for pattern recognition and understanding.

RESEARCH METHODOLOGY

A complete analysis makes sure to look at every possible danger and opportunity. Management and/or experts from inside and outside the company will shortly assess the outcomes. One or more organised methods may be used to arrange all of the analyses. To better convey risks, it is helpful to rank them according to their possible impact and source. The generation of reports and analyses vital to upper management relies on meticulous record-keeping and adherence to all regulations. Process visibility and understanding are both improved by this method. Strategy, finances, people, technology, and environmental hazards are just a few of the many that public sector organisations may encounter. Before you go ahead and start identifying risks, make sure you have a solid strategy for the report you'll make from the data. The research divides the company into sections according to the dangers it poses. Possible dangers include things like the loss of important employees, drastic cutbacks to funding, problems with communication, and physical infrastructure damage from natural catastrophes. Risk management measures must be put in place to protect the organization's goals against unanticipated problems. Finding and maintaining adequate levels of risk is a critical responsibility of those in authoritative positions. An organization's performance may be greatly affected by many hazards. To effectively manage these risks, one must first identify which risks have the most impact and then develop ways to reduce those risks.

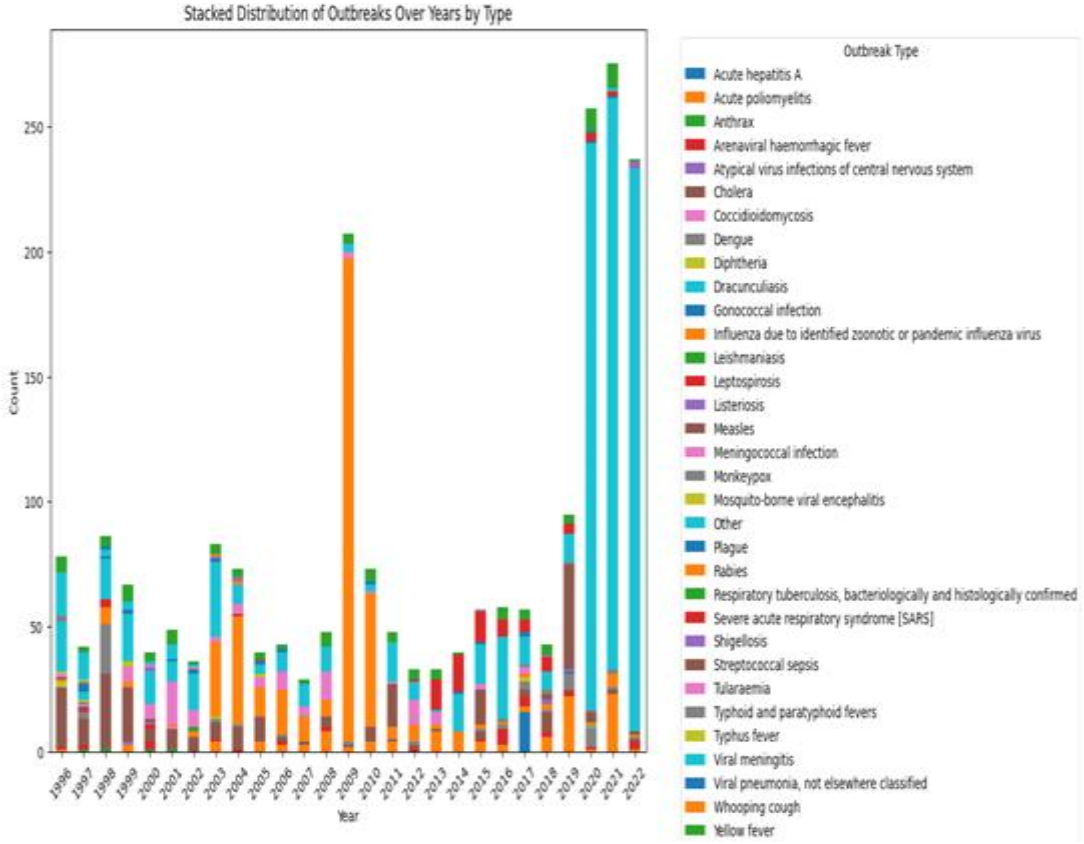
RESULTS AND DISCUSSIONS

The identification of COVID-19 and the assessment of mortality risk and severity are the main foci of our study. Clinical and laboratory data that is available to the public is used in this study. Algorithm creation, training datasets, and feature selection are some of the subjects addressed in this course.



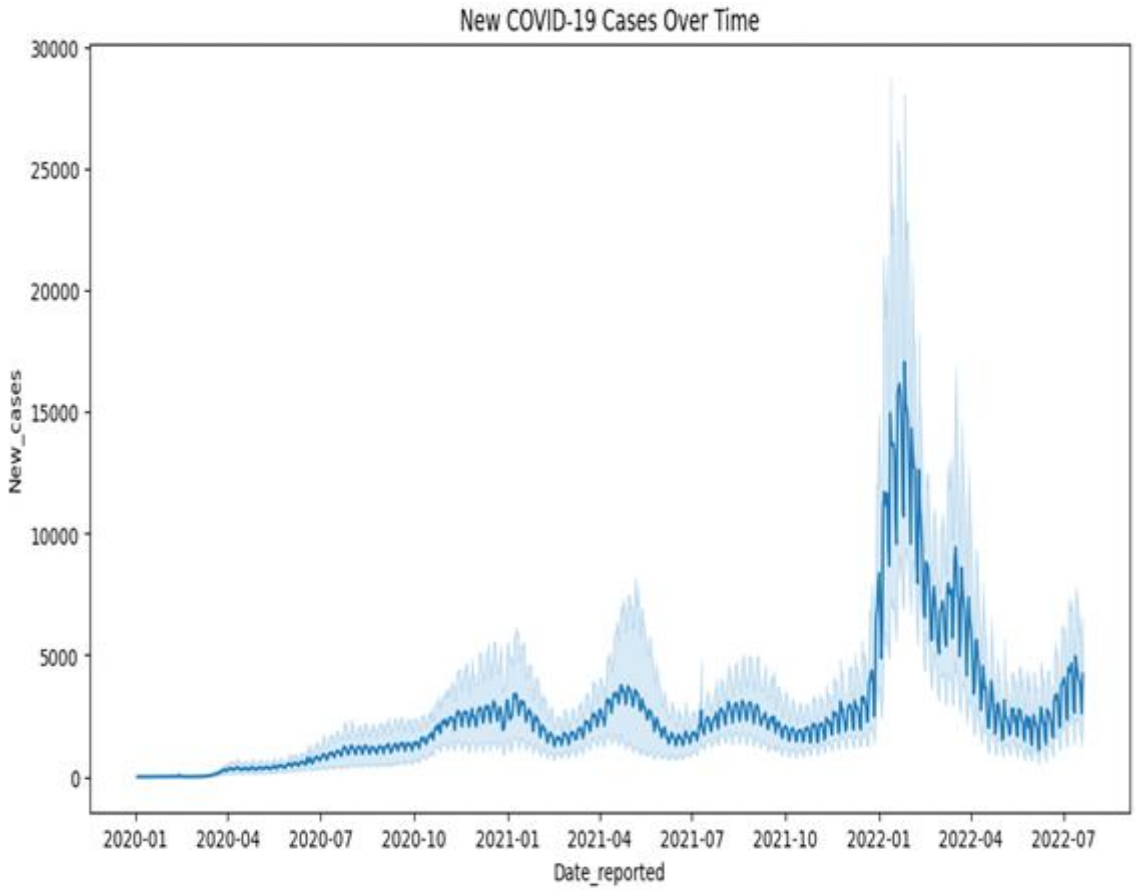
Graph 1: Distribution of Different Outbreaks over Years

This will help us understand the spread of diseases more thoroughly. Hybrid stacked LSTM-GRU optimised models, Prophet, and ARIMA are among the approaches that are used. The data frame that is part of the Prophet architecture makes dealing with data that is time series or has seasonality much easier. It is essential for a data frame to include these two columns.



Graph 2: Stacked Distribution of Outbreaks over Years by Type

In this column, which is labelled as "ds," the date-time series is shown. For the time series data frame, the relevant values are in the "y" column. For time series data analysis and prediction, the ARIMA models are generally considered the best statistical frameworks. This strategy, which comprises finding recurring patterns in the data, is a great way to get accurate time series predictions. The proposed model is an improved take on the original autoregressive moving average formula thanks to the use of integration. Our optimised hybrid model, which combines stacked LSTM-GRU layers, outperforms competing models when it comes to predicting the Covid-19 situation with little loss.



Graph 3: New COVID-19 Cases over Time

The model was optimised using the Adam optimiser and the ReLU activation function with a learning rate of 0.01. When compared to Prophet and ARIMA models, ours reduces loss significantly and performs better overall. Very little research has focused on hybrid models as a means of COVID-19 case predictions. The performance advantages of hybrid models over their separate components are well acknowledged. Consequently, looking into hybrid models to improve COVID-19 predictions is essential.

CONCLUSIONS

Machine learning algorithms are very valuable in public health for providing comprehensive and individualized risk assessments. These predictions have the potential to guide targeted therapies and improve overall outcomes. A thorough familiarity with computer systems enables people to keep a constant eye out for potential threats, which in turn allows for the prompt implementation of safeguards. Machine learning algorithms are used with real-time data sources, such as mobile health applications and Internet of Things (IoT) sensors, to accomplish this capacity. Organizations in the public sector may make better use of their resources if they use reliable risk assessments. This allows them to prioritize areas and people most likely to be impacted by health and safety issues, and to allocate resources accordingly. Despite the promising future of machine learning, questions of equality and bias in algorithms persist. To ensure that models reliably provide unbiased predictions, it is crucial to train them with data that accurately reflects the whole population. Machine learning models are very scalable, which is why they are cost-effective for risk management in several public

health domains. If public health machine learning models are to be effective and reliable, it is imperative that politicians, healthcare professionals, and data scientists work together.

REFERENCES

1. *Giovanna Maria Dimitri (2023), "A Comparison of Machine Learning Approaches for Predicting Employee Attrition", Appl. Sci., ISSNno: 2076-3417, Vol. 13(1), Pages. 267.https://doi.org/10.3390/app13010267*
2. *Subhash Monda [2023], "Machine learning-based maternal health risk prediction model for Io MT framework", International Journal of Experimental Research and Review, ISSN:2455-4855, Vol.32, PAGES. 145-159*
3. *Mohamed Zul Fadhli Khairuddin [2022], "Occupational Injury Risk Mitigation: Machine Learning Approach and Feature Optimization for Smart Workplace Surveillance", International journal of environmental research and public health, ISSN:1660-4601,vol.19,issue.(21)*
4. *Juan Lu [2021], "Machine learning risk prediction model for acute coronary syndrome and death from use of non-steroidal anti-inflammatory drugs in administrative data", Scientific Reports, ISSN: 2045-2322, vol.11, doi:10.1038/ s4159 8-021-97643-3*
5. *Li-Ya Wu [2021], "Ensemble Learning Models for Food Safety Risk Prediction", Sustainability, ISSN: 2071-1050, vol.13, issue. (21),https://doi.or g/10.3390/su13211 2291*
6. *Mangaiyarkarasi Thiagarajan [2021], "Credit Risk Modelling For Indian Debt Securities Using Machine Learning," Bulletin of Monetary Economics and Banking, ISSN:2460-9196,vol.24,pages.107-128*
7. *Maren E. Shipe [2019], "Developing prediction models for clinical use using logistic regression: an overview", Journal of Thoracic Disease, ISSN 2077-6624, Vol.11, doi:10.21037/jtd.2019.01.25*
8. *Md.Razu Ahmed [2020], "Breast Cancer Risk Prediction based on Six Machine Learning Algorithms", computer sciences and engineering, ISSN 2347-2693, vol.48, DOI: 10.1109/CSDE50874.2020.9411572*
9. *Megha Nain [2021], "Safety and Compliance Management System Using Computer Vision and Deep Learning", IOP Conf. Series: Materials Science and Engineering, ISSN:1757-899X,vol.1099,doi:10.1088/1757-899X/1099/ 1/0120 13*
10. *Min Chen [2017], "Disease Prediction by Machine Learning Over Big Data From Healthcare Communities", IEEE Access, ISSN: 2169-3536, Vol.5, DOI:10.1109/ ACCESS.2017.2694446*