Lab Report - 01

Experiment Name: Research Paper Review

Course Code: CSE326

Course Title: Data mining & machine learning Lab

Submitted To:

Name: Mr. Md. Firoz Hasan

Sr. Lecturer

Department of CSE

Daffodil International University

Submitted By:

Name: Md. Sohag Ali

ID: 0242220005101426

Section:63_B2

Department of CSE

Daffodil International University

Submission Date: 24-05-2025

1. **Title:** Dengue Dynamics in Bangladesh: Unveiling Insights Through Statistical and Machine Learning Analysis.

Abstract Summary	Dengue fever is a significant public health concern in Bangladesh with recurring outbreaks.1 The study uses statistical and machine learning analyses to identify temporal and spatial patterns, seasonal trends, and high-risk areas. It finds higher awareness levels among educated, semi-urban, female, and Dhaka residents. The epidemic is seasonal, peaking during the monsoon season, and is influenced by climate change and urbanization. The study advocates for targeted interventions and further research to improve prevention and control strategies, utilizing machine learning for forecasting and management.
Methodology	Primary and secondary data sources were used. Primary data was collected from dengue patients in hospitals using a pre-coded questionnaire. Data preprocessing was conducted for quality assurance. Statistical techniques included descriptive statistics, correlation analysis, and regression analysis. Prophet forecasting tool was used for machine learning analysis. Logistic regression was used to identify factors related to dengue awareness and knowledge.
Limitations	• Omitted participants due to incomplete responses. • Limited sample size. • Need for ongoing research to validate ML models and identify risk factors. • Further research needed for effective prevention and control strategies.
Future Scope	• Enhancing the accuracy of dengue prediction, forecasting, and control strategies in Bangladesh. • Offering evidence-based policy recommendations for governmental and healthcare authorities. - Implementing interventions such as enhanced surveillance, targeted vector control, public awareness campaigns, and improved healthcare access. • Conducting further research to develop more effective prevention and control strategies. • Utilizing machine learning analysis for forecasting and managing dengue outbreaks.
Data Collection	Data sources: Primary and secondary data . Primary data collection: From dengue patients in public and private hospitals in Dhaka city and other regions (Rajshahi, Barisal, Chittagong, Khulna, Mymensingh, Rangpur) . Secondary data source: Directorate General of Health Services (DGHS) in Bangladesh. Data collection tool: Pre-coded questionnaire. Sample size: 699 participants. Data preprocessing: Ensured cleanliness and suitability for analysis. Data sections: Socio-demographic characteristics and knowledge/awareness about dengue.

2. **Title:** Land use and meteorological influences on dengue transmission dynamics in Dhaka city, Bangladesh

Abstract Summary	Background: Dengue fever is a significant public health issue in Dhaka city, Bangladesh, spread by Aedes mosquitoes. The dengue indexes are highest in residential mixed zones. Environmental parameters like temperature, relative humidity, rainfall, and air pollution index are linked to mosquito larvae and dengue prevalence, influenced by urbanization and population expansion. Findings: Land use significantly influences dengue incidence patterns. There is a significant correlation between climatic characteristics (temperature, rainfall, specific humidity, relative humidity, wind speed) and dengue incidence patterns. Implications: The study demonstrates a complex relationship between land use, meteorological factors, and dengue spread. Climatic conditions play a significant role in driving dengue epidemics. The results have implications for urban planning, public health measures, and vector control tactics.
Methodology	Spatial and temporal analytic methodologies are used to investigate correlations between land use attributes, climatic variables, and dengue fever occurrence. ArcPy and ArcMap 10.7 are used for geographic data analysis and visualization. A quantitative approach is taken, examining meteorological and land use data. Numpy and Pandas are used for data extraction and analysis.4 - Microsoft Excel is used for data visualization. Jupyter Notebook and IDLE are used for scripting. The Pearson correlation coefficient is used to quantify linear associations.
Limitations	• Significant gaps and incompleteness in the dataset. • Structural problems with the monitoring system. • Lack of awareness among responsible authorities.
Future Scope	• The future scope of this study includes informing urban planning, public health measures, and vector control tactics. • It also involves developing a climate and land use-based early warning system for dengue outbreaks and incorporating health-related factors into land use planning to address public health needs.
Data Collection	Dengue data: Collected from the Directorate General of Health Services (DGHS) website, including daily numbers of dengue incidents and deaths. Meteorological data: Obtained from the POWER Project of NASA,2 including time-series data from September 2022 to August 2023. Dataset: Consists of observations over 350 days. Data availability: Will be provided upon request.

3. **Title:** Prediction of Dengue using Machine Learning Algorithms: Case Study Dhaka

Abstract Summary	Bangladesh and South Asia face a significant problem with dengue fever. Global dengue incidence has grown dramatically, with 100-400 million infections and 22,000 deaths annually. The study aims to predict dengue outbreaks using environmental factors like rainfall, humidity, and temperature. The proposed machine learning model uses Support Vector Machine (SVM) for prediction. The SVM model achieved 97% accuracy in predicting dengue epidemics in Bangladesh. The study contributes an efficient machine learning model with limited data for predicting dengue.
Methodology	Used three machine learning algorithms: SVM, XGBOOST, and logistic regression. Classified outputs into "epidemic" and "non-epidemic" classes. Implemented models using Python's Jupyter notebook. Conducted parameter tuning using randomized search. Preprocessed data with stratified train-test splitting and SMOTE for imbalance handling. Normalized data using MinMax scaler. Compared model performances to determine the best model.
Limitations	• Limited availability of data on dengue cases in Bangladesh • Need for a larger dataset with more predictors for improved accuracy
Future Scope	• The future scope involves creating a large dataset of Bangladesh with more predictors to get a strong model and improve accuracy.
Data Collection	Data source: Published research paper titled "Long-term predictors of dengue outbreaks in Bangladesh: A data mining approach." Dataset size: 240 entries with 11 columns each. Data period: Monthly data from 2000 to 2009. Data attributes: Month number, rainfall, humidity, temperature, dengue cases for two cities. Preprocessing: Selected relevant data, removed non-Dhaka city data, added epidemic indicator. Data splitting: Stratified train-test split (60% training, 20% testing, 20% unknown sample). Balancing: SMOTE technique used to balance class imbalance.

4. **Title:** Smart system for real time monitoring and diagnosis of dengue surfaces in Bangladesh

Abstract Summary	Efficient vector management is crucial for reducing vector-borne diseases. UAVs can be used to detect mosquito breeding sites. A methodology using a CBAM-enhanced YOLOv9 framework is proposed for automatic detection in aerial images. The system achieves high performance metrics: mAP50 of 99.5%, mAP50-95 of 86.4%, IoU of 94%, 45 FPS, 98% precision, and 90% recall. It is robust across urban and rural environments and facilitates timely vector control interventions.
Methodology	UAV-based system for monitoring dengue surfaces CBAM-enhanced YOLOv9 object detection framework. Manual and automated data collection using UAV-mounted cameras. Object-level annotation using LabelImg. Image preprocessing, data annotation, YOLO model training, and evaluation.
Limitations	• Extensive labor and time required for manual image inspection. • Need for further testing under various circumstances. • Challenges with UAV-based monitoring due to wind and lighting issues. • Regulatory restrictions on UAV operations. • Future research directions include integrating multispectral and hyperspectral sensors, real-time edge AI systems, differential privacy techniques, and expanding the dataset.
Future Scope	• Confirming the model's practicality through more tests under various circumstances. • Optimizing data gathering and analysis with AI-based image processing technologies. • Integrating multispectral and hyperspectral sensors with UAV systems to improve detection performance. • Deploying real-time edge AI systems on UAVs for on-the-fly mosquito breeding site identification and treatment. • Investigating differential privacy techniques and federated learning models for data security and ethical concerns. • Expanding the dataset to cover diverse geographic regions and seasons for improved robustness and generalizability.
Data Collection	Data collection involved both manual sourcing and expansion using UAV-mounted cameras. The dataset consists of 1500 images, each with dimensions of 800 × 600 pixels. Images were obtained from urban and semi-urban regions in Dhaka and Gazipur districts, Bangladesh. Images feature containers that could harbor mosquito larvae, such as plastic buckets and tires. Data was collected using a Pixhawk flight controller and FPV camera. UAV operated at heights between 10-50 meters for detailed observation. Data collection followed UAV regulations and maintained privacy standards.

5. **Title:** A Conceptual Framework of an Automated Mosquito Control in Drainage Systems for Combating Dengue in Bangladesh

Abstract Summary	Dengue fever is causing widespread suffering in Bangladesh, with increasing deaths and hospitalizations. The Aedes mosquito is the primary vector of Dengue. Traditional fogging methods are ineffective and harmful. The paper proposes an automated, cost-effective, and efficient mosquito control system for drainage systems. The system has a detection success rate of over 80%.
Methodology	Use of Raspberry Pi as the central processor. Incorporation of Pi camera for mosquito detection. Use of a touch sensor with a luring pad to attract and trap mosquitoes. Employment of a laser to kill mosquitoes. Training and deployment of a neural network model for detection and decision-making. Evaluation of the neural network model's performance using accuracy, precision, recall, and f1-score.
Limitations	• The system is static and needs to be attached to something. • The system can only detect mosquitoes, not specifically those that are carriers of Dengue. • Memory limitation of Raspberry Pi prevents the use of more advanced neural network models. • Future work needed to make the system mobile and integrate more powerful processors and neural network models.
Future Scope	• Future work will focus on the full-fledged implementation of the system, making it movable and using more powerful processors and neural network models at affordable cost, addressing current limitations such as static nature and memory constraints of Raspberry Pi.
Data Collection	Dataset source: Open-access dataset from Kaggle Dataset size: 1200 images (600 Aedes mosquitoes + 600 "Not Mosquito" images1) Data split: 80% for training, 20% for testing2 - Data augmentation: Performed to create class balance and avoid bias. Image preprocessing: Images cropped to region of interest. Model training: Trained on a computer, deployed on Raspberry Pi. Performance evaluation: Accuracy6, precision, recall, F1-score

6. **Title:** Forecasting Dengue Incidence Rate in Tamil Nadu Using ARIMA Time Series Model

Abstract Summary	Mosquitoes are a significant health threat in India, spreading diseases like dengue fever. The incidence of dengue has increased 30-fold globally over the past 30 years. The spread of dengue in India is attributed to factors like climate change and urbanization. The paper proposes using an ARIMA time-series model to forecast dengue incidence rates in Tamil Nadu. The model uses meteorological data to predict dengue cases, providing a tool for health officials to take preventive measures.
Methodology	ARIMA time-series model for forecasting dengue incidence rate using meteorological data. Preprocessing: plotting data, removing outliers and missing values. Decomposing data into trend, seasonal, and random components. Fitting ARIMA model using historical empirical data and autocorrelation analysis. Enhancing prediction by adding seasonal variables and refitting the model.
Limitations	• Current monitoring systems in Tamil Nadu face challenges such as preference for eradication over surveillance, difficulty in interpreting findings, and lack of coordination between units. • The trust boundaries of the forecast indicate a need for a more reliable model. • The forecast model is described as simplistic, suggesting it may not capture all complexities of dengue incidence rates.
Future Scope	• The future scope involves improving current methods of tracking and forecasting mosquito-borne diseases, particularly through the use of ARIMA models for predicting dengue cases. • There is a need for more reliable models to enhance predictive accuracy and support timely preventive measures.
Data Collection	Two datasets were used to predict the dengue incidence rate in Tamil Nadu.1 - Data included historical dengue case data and meteorological data (TMAX, TMIN, TAVG, RAINFALL, PRCP). Time series data was collected for forecasting purposes using ARIMA models.

7. Title: Mosquito Miner: A Light Weight Rover for Detecting and Eliminating Mosquito Breeding Sites

Abstract Summary	In this paper, we present a novel approach to the development and deployment of an autonomous mosquito breeding place detector rover with the object and obstacle detection capabilities to control mosquitoes. Mosquito-borne diseases continue to pose significant health threats globally, with conventional control methods proving slow and inefficient. To mitigate the limitations of manual labor and traditional methods, our rover employs autonomous control strategies. Leveraging our own custom dataset, the rover can autonomously navigate along a pre-defined path, identifying and mitigating potential breeding grounds with precision. It then proceeds to eliminate these breeding grounds by spraying a chemical agent, effectively eradicating mosquito habitats. Our project demonstrates the effectiveness that is absent in traditional ways of controlling and safeguarding public health.
Methodology	Development of an autonomous rover, Mosquito Miner, for detecting and eliminating mosquito breeding sites. Use of a custom dataset to train a machine learning model for identifying mosquito hotspots. Navigation along a pre-selected path using GPS and obstacle detection systems. Image recognition algorithms for identifying potential breeding sites. Spraying a chemical agent to eliminate breeding sites. Performance assessment using Mean Average Precision (mAP@50) and area coverage metrics.
Limitations	• Room for improvement in detecting all potential threats. • Limited processing power and efficiency of the current hardware. • Limited comprehensiveness of the current dataset. • Limited structural capabilities for navigation and operation in challenging environments.
Future Scope	• Upgrade the rover's hardware by replacing Raspberry Pi3 with a more sophisticated microcontroller to enhance processing power, efficiency, and robust performance. • Enhance the dataset with more compatible data to improve recognition of different mosquito breeding patterns. • Structurally enhance the rover to navigate and operate effectively in challenging environments.
Data Collection	The authors used a custom dataset for training the machine learning model to identify mosquito breeding sites. The details of how this dataset was collected are not provided in the paper.

8. Title: Observing the Unobserved: A Newspaper Based Dengue Surveillance System for the Low-Income Regions of Bangladesh

Abstract Summary	Dengue is a significant health and economic burden in middle to lower-income countries due to its high morbidity and mortality rates. Bangladesh, with its history of dengue outbreaks and high medical expenditure, lacks a stable surveillance system. This paper uses data mining techniques on local Bengali newspapers to develop a dengue surveillance system, achieving an f-score of 91.45% and identifying under-reported regions while linking socio-economic factors to dengue reporting.
Methodology	Data-mining techniques were used on local Bengali newspapers to develop a dengue surveillance system.1 - A corpus of dengue-related news was developed from 270 online newspapers from 2017 to 2019. Data preprocessing involved removing stop words, URLs, and special characters. News articles were categorized into intervention and disease classes. Seed-guided topic modeling was used to expand keyword sets. Machine learning models (KNN, Multinomial Naive Bayes, SVM) were trained and evaluated for classification. The methodology was applied to a large corpus of news articles to explore socioeconomic and demographic factors.
Limitations	• Inability to extend observations beyond three years due to unreliable news archive data before 2017. • Lack of access to all official government records, leading to reliance on third-party sources.
Future Scope	• The future scope involves improving the newspaper based surveillance system by extending observations beyond three years, accessing more official government records, and leveraging increasing data quality to enhance the system's effectiveness and applicability for monitoring dengue and other endemic diseases in Bangladesh.
Data Collection	Collected dengue-related news from 270 online Bengali newspapers from 2017 to 2019. Used a specialized web crawler from the Bengali Search Engine, Pipilika, to collect archived daily news. Collected dengue-related records from Access to Information Programme (a2i)3, CDC, IEDCR, and DGHS. Performed time series analysis of officially tracked data and newspaper archives. Collected additional data from Bangladesh Bureau of Statistics (BBS) yearbook and The World Bank.