PASIGLA: A Smart Healthcare Platform for Pasig Local Health Centers Featuring Machine Learning-Driven Disease Prediction and Optimized Medication Inventory Management

A Thesis Project proposal presented to the

College of Computer Studies of

Pamantasan ng Lungsod ng Pasig

In partial fulfillment of the requirement for the degree of

Bachelor of Science in Computer Science

by:

De Justo, Izzy O.

Lausa, Daniel P.

Pili, Krisna Raya S.

Noreen A. Perez, DIT

Adviser

June 2025

Table of Contents

1.1 The Problem and Its Background

Community health centers serve the critical function of offering accessible care and responding to local health needs. Yet many healthcare facilities in the Philippines, especially at the city and barangay levels, experience operational issues like inadequate disease surveillance and stock management of medicines. These may result in late response to disease outbreaks and either a shortage of the necessary medicines or a surplus.

Over the last decade, it has become increasingly challenging for public health systems to identify both communicable and non-communicable diseases early and manage them appropriately. Diabetes and dengue are still major health concerns in the Philippines, reaching thousands of individuals annually. The COVID-19 pandemic, which reached its peak between 2020 and 2021, also highlighted the need for active health surveillance and data-driven predictive technologies. This crisis exposed the need for smarter healthcare systems that can anticipate public health demand and respond accordingly.

Besides, the Philippines has substantial primary healthcare gaps owing to disintegrated service delivery, insufficient health financing, and an ill-motivated workforce. These are the reasons why the enactment of the Universal Health Care (UHC) Law was prompted, which would solve these problems by providing universal coverage for quality care, promoting health equities, and providing a broad primary care benefit package. The UHC law also provides changes in PhilHealth, encourages integrated services among local government units, and creates the Special Health Fund to facilitate better resource allocation. In spite of continued implementation issues, these reforms are directed towards equitable access and better health outcomes.

Pasig City, which is among the emerging urban cities in Metro Manila, has significant potential for digital progress in healthcare. Aside from the data from the Department of Health and local sources, there is still a wide gap in the use of such data to bring about useful information. Intelligent technologies, such as machine learning, are also not extensively utilized to predict health trends or streamline important operations such as medication inventory control.

To meet these gaps, this research introduces PASIGLA, a Smart Healthcare Platform for the Pasig community health facilities. The platform will have machine learning built-in to determine the potential occurrence of prevalent diseases, especially dengue and diabetes, at the barangay level. By taking advantage of publicly available health data between 2014 and 2024, with a focus on the pandemic years of 2020–2021, PASIGLA hopes to deliver timely information useful in enhancing preventive strategies and optimum resource allocation by healthcare providers and local governments.

1.2 Review of Related Literature

Utilizing the power of different algorithms, modeling, and artificial intelligence, evaluating historical data of a patient or even diseases become easier predicting the likelihood of future illnesses, outbreaks, and necessary medical interventions with greater accuracy and speed. A model was developed by (Gao et al., 2020) to predict the mortality risk of a patient during the COVID-19 pandemic that “uses clinical data from patients' admissions to stratify them by mortality risk, allowing for the prediction of physiological deterioration and death up to 20 days ahead.” This accurately predicts risk stratification of COVID-19 patients upon admission. Helping clinicians to identify patients with mild to severe symptoms of the disease. Another study conducted by (Onay & Onay, 2019) applied machine learning in finding drug molecules to potentially find that might work against a specific disease. Machine learning accelerates this process by learning the patterns from past data about the drugs such as which ones worked, did not, and for which disease. The researchers developed a Drug Decision Support System (DDSS) to categorize each drug candidate molecule as possibly drug or non-drug, and to predict its illness class. This classification model serves as a filtration system in drug creation to eliminate potential inappropriate molecules in the early stages. Researchers calculated the chemical properties of a drug by using Molecular Descriptors which assign numerical values to describe properties of drug molecules, this will help them decide whether a molecule will act like a drug. Utilizing Lipinski's Rule of Five, researchers can make a judgment whether the drug can be taken orally or not by checking its molecular weight, number of hydrogen bond acceptors, and solubility and permeability. They found common molecular patterns which helped them identify common structures in drugs that were withdrawn or removed from the market due to problems. The ANN performed very well in correctly identifying and classifying the drugs with 84.6 percent and 83.3 percent accuracies on test sets including various application of drugs such as cardiac therapy, anti-epileptic, and anti-Parkinson's drugs.

An overview analysis of (Uddin et al., 2019) comparing various supervised machine learning algorithms from numerous studies published in two databases Scopus and PubMed where 48 articles were selected for the comparison. Results found that among various machine learning algorithms, Support Vector Machine (SVM), Naïve Bayes algorithm, and Random Forest are the most used. In accuracy, Random Forest showed the highest in 9 of the studies over 17, around 53 percent. Followed by SVM with 41 percent of the studies it was considered. Study conducted by (Venkatesh & Raju, 2023) utilizing Machine Learning algorithms, deep learning, and Streamlit. Researchers utilized the Support Vector Machine algorithm, TensorFlow with Keras, and Logistic Regression to predict five disease categories namely Diabetes, Heart Disease, Parkinson's Disease, Kidney Disease, and Breast Cancer. Comparing existing system accuracy to the system the researchers developed, the results showed the created system is much more accurate than the existing ones. In predicting diabetes using the SVM classifier algorithm, the existing system accuracy is 76 percent while the proposed system accuracy is 78 percent. In heart disease using Logistic Regression, 80 percent in existing ones and 85 percent on the proposed system. In Parkinson's disease using SVM classifier, 71 percent accurate on the existing system and 87 percent on the proposed system. As mentioned, the two additional diseases kidney disease and Breast Cancer do not have an existing system leaving nothing to compare with, yet it performed very well with 97 and 96 percent of accuracy respectively.

Inventory management is the process of ordering, storing, using, and selling a company's inventory. This covers raw materials, components, and completed goods, as well as the storage and processing of these items. There are various techniques of inventory management, each with advantages and disadvantages based on a company's requirements as defined by (Hayes, 2024). In the context of public healthcare, it refers to the effective control and optimization of essential supplies and resources within healthcare facilities which includes management of drugs, medical equipment, personal protective equipment (PPE), and other materials needed for patient care and public health programs. In the Philippines setting, numerous problems arise challenging proper inventory management in various public healthcare facilities. Reported by (Gonzalez, 2023) in Rappler that inadequate procurement planning and poor distribution and monitoring mechanisms resulted in the waste of over billions of pesos worth of pharmaceuticals and medicines in the Department of Health (DOH) inventory. With over 7.431billion pesos worth of medicines and drugs were expired, nearly expiry, damaged, overstocked, excessive, low usage, and undistributed or being distributed late as labeled by the Commission on Audit (COA). These medicines and drugs should have been distributed to various locations such as barangay level public health offices and pharmacies available for the public to use and utilize. With the advent of technology, as one of the subcategories of AI or artificial intelligence, Machine Learning can be utilized in inventory management. Machine Learning can help in forecasting future usage or even sales by analyzing past usage of various medicines or drugs allowing strategic use and handouts of medicines ready to be distributed to the public and the ideal stock level for each medicine and drugs. As per (Dorota-Owczarek, 2024), reducing the risk of stock outages and excess inventory, promoting efficient supply chain operations, facilitating real-time decisions, and assisting businesses (public healthcare facilities in this context) in adapting to market shifts are the benefits of utilizing machine learning. Assisting them to properly create decisions. A study conducted by (Zwaida et al., 2021) aiming to improve the inventory management in a hospital and community-based system by introducing a “Deep Reinforcement Learning”. In comparison with the commonly-used method which is overprovisioning, the system presents better performance in terms of refilling cost. “In most timeslots, the over-provisioning approach is more expensive than the DRLD.” (Zwaida et al., 2021), indicating that the machine learning model effectively minimizes restocking expenses by learning optimal refill patterns and adapting to demand fluctuations. The DRLD system, on average, reduces the cost by 12.31% compared to over-provisioning method. IN terms of preventing shortages, the study proves that the system introduced performed best. With a shortage rate of 1.7% it demonstrates superior responsiveness and efficiency in preventing shortages compared to other methods, these methods are over-provisioning, ski-rental, and max-min. In terms of unexpected refilling or unnecessary restocking, the DRLD method showed the lowest average rate of 1.03%, indicating better stability and responsiveness in inventory management compared to over-provisioning with 1.675%, and ski-rental with 1.54%.

According to the (World Health Organization: WHO, 2019), Primary Health Care or PHC is where “Everyone has the right to the best possible health.” which is applied in our local community and seeks to address all health problems. In the growing world and a turn to technological advancement, the application of it was seen to be beneficial at a tremendous level. Creating accurate and faster diagnoses, efficient workspace for medical practitioners, and safer place for patients. Studies conducted by (Gao et al., 2020) and (Onay & Onay, 2019) confirm the application of machine learning in the medical field for the betterment of the patient and the general people. Development of systems detecting early symptoms of different diseases or illnesses contribute to overall health safety of the community, treating diseases as early as they emerge, preventing spread and contamination. By detecting molecular cues in the drug molecules and identifying whether this drug will work, will not work, and on which disease it is most effective will accelerate the creation of effective medication. Application of technology is the advent of comfort and safety is utilized properly.

Detection of disease trends and classification of a population are essential in planning health-related programs. Preventing a rise in cases by devising a plan accommodating the needed preemption in the spread of a certain disease in this time of the year. Studies utilizing machine learning greatly increase in the prevention of numerous illnesses. In the analysis by (Uddin et al., 2019), machine learning algorithms such as Support Vector Machine (SVM) where 48 articles were seen on usage of this machine learning algorithm. This is proven by (Venkatesh & Raju, 2023) where Support Vector Machine is one of the machine learning algorithms used in their study about drug identification. The analysis by (Uddin et al., 2019) is helpful for this study as this limits the scope of effective machine learning algorithms for this field of study.

Proper and efficient inventory management is important for local or community healthcare centers as they are continuously supplied with medication which will be distributed to the concerned people. Stocks, storage, and even prices are affecting matters on the supplies and demand on these aspects, so it is crucial on proper balancing and identifying factors that will affect the distribution and supply of the medication. Utilization of machine learning algorithms has been proven to be effective as demonstrated in the study of (Zwaida et al., 2021). In relation to this study, we can reference the methodology used by the researchers on developing the system or follow the procedures and formulation done to come up with proper results.

However, usage of such technology is not or merely accessible to local perspective. Facing multiple problems and strains such as limited infrastructure, connectivity issues, and access to digital health technology is hard. Various places in the Philippines struggle with these technological features due to inability to acquire one and have limited resources, hindering access to the benefits of such technology to reach patients. Additionally, despite the benefits of machine learning application in healthcare, it gains a lot of skepticism for its usage and ethical application.

1.3 Significance of the Study

Through the development of PASIGLA, a smart platform intended for Pasig City health centers, this study ran to deal with urgent issues in the administration of healthcare locally. The project provided a data-driven approach to improve public health services through strengthening medication inventory systems and integrating machine learning for disease prediction. The study mattered in more ways than one, proving that even simple use of technology can make a real difference—not just for one place, but for others who face the same challenges.

Global Context. PASIGLA was built in response to the increasing need for practical, intelligent, and accessible healthcare solutions around the world. As more countries began exploring how AI and data analytics could be used in public health, this study became a simple yet powerful example of how technology could help detect diseases earlier and manage medicine supplies more effectively. It demonstrated that, even with inadequate funding, developing countries could innovate while achieving significant modifications to their healthcare systems.

Economical Context. Budget limitations have long been a challenge for public health centers. This study addressed that by introducing a simple system that helped staff keep track of which medicines were needed the most and when. With this, health centers were able to avoid wasting supplies or running out of important medications, making the most out of their limited resources.

Environmental Context. This study helped promote mindful and sustainable practices in healthcare. By making it easier for health workers to monitor medicine stocks, it reduced the chances of medicines expiring and going to waste. It also encouraged health centers to be more responsible in how they handled medical supplies, lessening their environmental footprint in the process.

Societal Context. The PASIGLA system benefited individuals by:

- Making it easier for health centers to respond quickly and properly to patients, since it was built based on what the community really needed.
- Bringing back people's trust in local health centers by making sure services and medicines were ready when needed, which encouraged more residents to visit and ask for help.
- Supporting Pasig City's effort to build a healthier and better-informed community by helping people understand their health needs through clear and timely information.

Researchers and Future Research. Aside from addressing a specific local concern, this study aimed to encourage further exploration in the field. It opened new opportunities for professionals, scholars, and students who are interested in how technology can be used to improve public services. PASIGLA showed that even small, practical innovations can lead to bigger projects in areas such as healthcare, disaster response, and other public sectors. Future researchers may continue this initiative by improving its features, expanding its scope, or adapting it to fit different communities and settings.

1.4 Statement of the Problem

Most Health centers in Pasig face such challenges in how they carefully manage the medicine supplies and detecting illnesses that are high or on trend. As a result, there are times when patients do not receive timely care. This is a bigger problem in areas where people rely mainly on public health services for treatment.

Many health centers still rely on manual systems to track medicine stocks and monitor patient health concerns. These outdated processes make it easy for errors to happen — such as running out of important medicines, keeping too much of others, or missing early signs of common illnesses. Without the right tools or systems, health workers find it difficult to provide care quickly and correctly. This study identified the need for a system that could help forecast potential health issues and keep track of medicine stocks more effectively, making it easier to respond when needed.

The goal of this study was to build PASIGLA, a smart healthcare platform made for Pasig's local health centers. It was designed to help health workers detect common illnesses earlier using machine learning and to organize their medicine inventory more effectively. Through this system, health centers could avoid overstocking or shortages and become more prepared in giving the right care at the right time. This aim is to create a smoother and more responsive service for the people in the community.

1.5 Scope and Limitation of the Study

The software focuses on two features: machine learning-based illness prediction and pharmaceutical inventory optimization. It is meant to provide health professionals and local governments with predictive information about the likelihood of prevalent illnesses such as dengue and diabetes occurring at the barangay level, based on publicly available health data. It also aims to improve the accuracy and efficiency of controlling drug inventories to avoid shortages and overstocking.

The study also aims to use machine learning techniques using Python to analyze historical health data from 2014 to 2024, with an emphasis on the 2020-2021 pandemic year. The algorithms will be used to anticipate illness incidence and allocate medical supplies more efficiently. The integration of inventory monitoring technologies will allow for real-time monitoring and forecasting of drug requirements based on health patterns. The system also aims to have an intuitive interface that is easy to browse for health personnel and administrative staff, hence boosting its usability and functionality.

The validity and reliability of the system mostly rely on the quality and accuracy of the health information obtained. Erroneous, incomplete, or inconsistent records may considerably impact the predictions made by the machine learning algorithms. The system will also be narrowed in scope to the diseases and barangays to which adequate data exist, and predictions will not be made on an individual patient basis to help maintain privacy. Moreover, developing and running such a system would need to rely on high-level hardware, constant software upgrades, and technical capabilities, which consume resources. The representativeness and diversity of the training dataset itself also contribute a significant share in making predictions accurate, unrepresentative or biased data can result in misleading outcomes or under-predicted risks of disease.

2.1 Introduction

This chapter introduces the methodology employed by the researchers in developing the PASIGLA system. Outlining the algorithm used for disease and stock predictions, the programming languages and tools utilized in building the platform and dashboard, and the system framework that defines its core functions and features.

2.2 Research Design

The study uses a mixed-methods research strategy, combining quantitative and qualitative techniques to fully meet the PASIGLA system's goals. It is mostly quantitative since machine learning models are trained and assessed using numerical data, such as pharmaceutical inventory records and illness case counts by barangay. Response planning and resource allocation can be optimized by using the predictions these models produce about the probability of illness occurrence. The research is also experimental. Various machine learning algorithms are tested and compared based on their performance metrics. The goal of determining which algorithm is best suited for forecasting common disease using the available data from 2014 to 2024.

2.2.1 Architectural Framework

Developing a system that involves data processing, user interactions, and interface display is well-suited to the Model-View-Controller (MVC) architectural framework. Defined by (MVC - Glossary | MDN, 2023) this design pattern separates the application into three interconnected components. Model, which handles the data and logic. View, which manages the user interface. And Controller, which acts as an intermediary between the Model and the View. Promoting modularity, maintainability, and scalability in system development. See the model below:

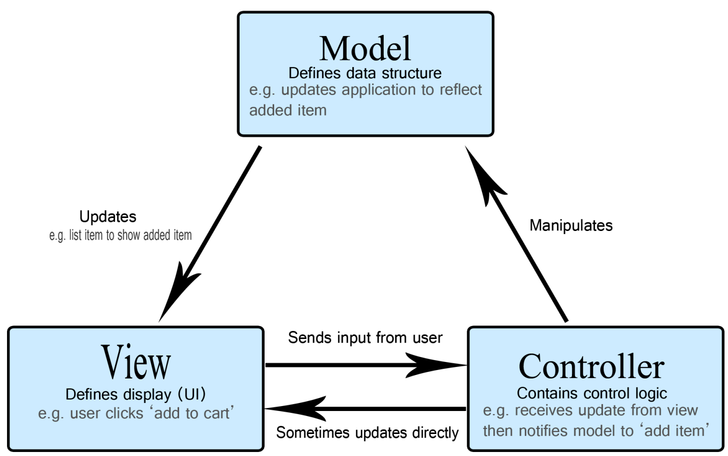

Figure 1: MVC Framework

The Model, which manages the data and business logic. In PASIGLA, this includes database tables that store values such as id, barangay_name, disease_name, disease_cases, and other relevant factors, including weather variables that affect disease transmission (e.g., rainfall and humidity for dengue). The view is the user interface allowing different roles to interact with the system and the data. The system contains a dashboard where the input of data, number of cases, stock number, alerts, and visualization of the disease trends, predictions, and barangay risk level. It is connected to the controller where the various logic are stored. This updates the view and/or the model when changes in the logic happen. Functions such as connection of each view, functions, input and output, and some features.

2.2.2 Algorithm

A range of machine learning algorithms were tested to determine the most suitable models for the PASIGLA system's predictive components. The evaluation focused on two primary tasks: disease prediction and inventory/stock forecasting. The algorithms tested include:

- Logistic Regression: It is a baseline classification algorithm known for its simplicity and interpretability which is ideal for the classification applied in this study which identifies whether a barangay is high-risk or low-risk elves.
- Naive Bayes: Effective in probabilistic classification problems where input features are independent. It offers fast computation and performs well even with relatively small datasets.

- Support Vector Machine (SVM): is known to perform well in high-dimensional spaces and handle non-linear decision boundaries. SVM is used to classify complex patterns in disease risk levels.
- Seasonal Autoregressive Integrated Moving Average (SARIMA): a time series model that incorporates trend and seasonality, making it suitable for forecasting disease cases and inventory patterns that follow seasonal fluctuations.
- Long Short-Term Memory (LSTM): A recurrent neural network (RNN) capable of learning long-term dependencies in sequential data. It is suitable for time series forecasting involving disease trends over time.
- Random Forest: Based on decision trees which can handle both classification and regression tasks. It is robust to overfitting and useful for modeling complex, non-linear relationships in healthcare and inventory management.
- XGBoost: A gradient boosting framework that performs well and is scalable. It is widely used in structured data prediction and offers superior accuracy through boosting techniques.
- K-Nearest Neighbors (K-NN): A non-parametric algorithm that predicts outcomes based on the closest training examples. It is intuitive and effective for pattern recognition in classification and regression tasks where the structure of historical data plays a critical role.

2.2.2.a Disease Prediction

For this task, the dataset used for disease prediction is about Dengue Cases in Pasig City per Barangay from 2018 to 2023. Columns are Year, Month, Barangay, Disease_cases, temperature, dew point, precipitation, and humidity. The last four columns were added to have a better analysis of the machine learning algorithms.

The algorithms are divided into two categories:

a. Regression type (dengue disease cases prediction)
    - SARIMA
    - LSTM
    - Random Forest
    - XGBoost
    - K-NN
b. Classification type (identifying high-risk and low-risk barangays)
    - Logistic Regression
    - Naive Bayes
    - SVM
    - Random Forest
    - XGBoost
    - K-NN

2.2.2.b Inventory/Stock Management

The PASIGLA system includes a predictive component for inventory forecasting. This is important in avoiding shortages or even overstocking of medicines, particularly in response to fluctuating disease cases. Time series prediction models were applied to forecast monthly medication demand based on past usage patterns and seasonal disease trends.

The forecasting models tested for inventory prediction include:

- Seasonal Autoregressive Integrated Moving Average (SARIMA)
- Long Short-Term Memory (LSTM)

These algorithms were trained using a sample dataset from Kaggle with columns of Date, Expense_Category, Amount, and Description. They are evaluated using the following evaluation metrics:

- Mean Absolute Error (MAE)
- Root Mean Square Error (RMSE)
- Mean Absolute Percentage Error (MAPE)
- R^2 Score

2.2.2 SE Paradigm

This section illustrates the connection between the components of the PASIGLA system, its function, and the flow of data. These visualizations will serve as a guide both the implementation and future maintenance of the system.

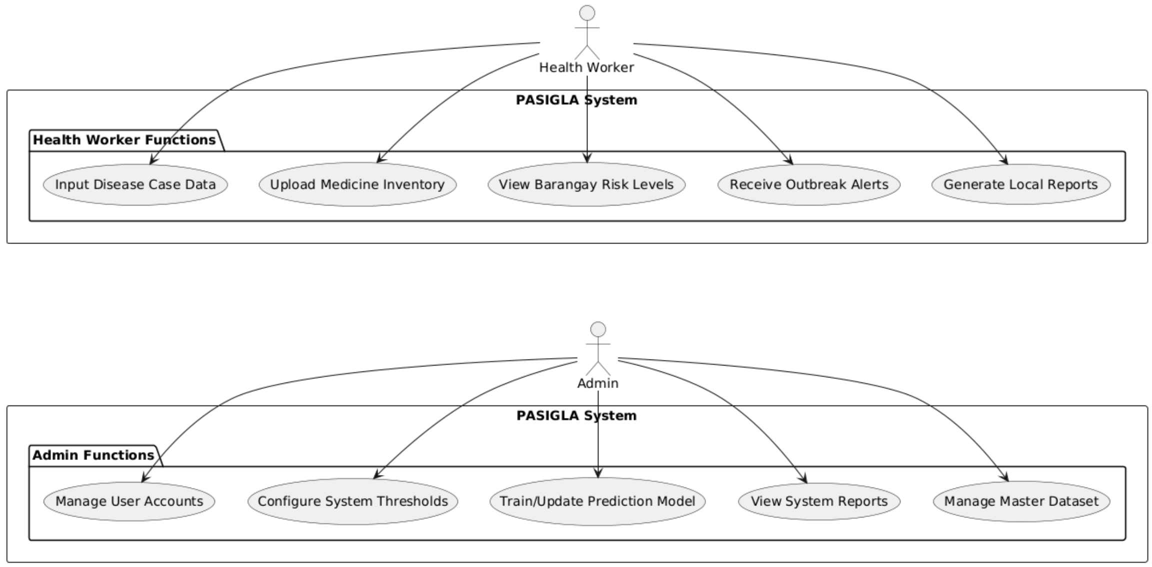

Figure 2: Use Case Diagram

Figure 2 presents the interactions of different roles or users in the system, as well as the functions they are allowed to use. Figure 2.a shows what an actor Health Worker can do and can access. Health workers are designated to view each barangay risk level and receive outbreak alerts, allowing planning and implementation of timely preventive measures, efficient resource allocation, and targeted health interventions within their assigned areas. Additionally, Health workers are also responsible for uploading and entering essential information, such as disease case data and medicine inventory, as they are the first point of contact in handling and recording these details. Lastly, generating reports for the current disease cases and inventory. Figure 2.b illustrates the administrative functions within the system, including the management of user accounts, disease case datasets, and medicine stock records. It also shows the admin's ability to configure system settings, train and update the prediction models, and view comprehensive system reports.

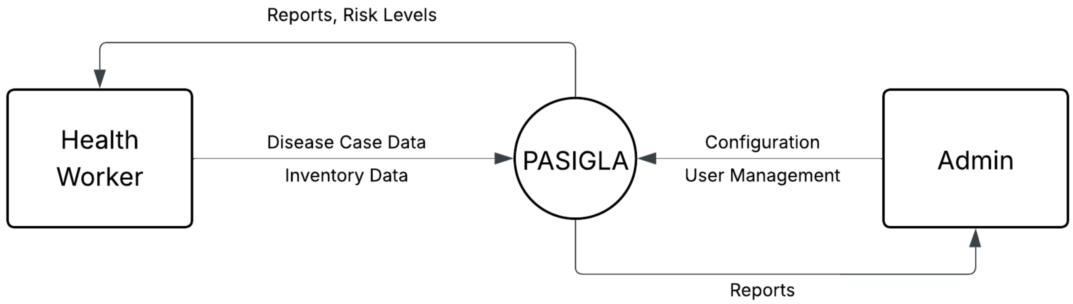

Figure 3: Level 0 Data Flow Diagram

Figure 3 presents the Level 0 Data Flow Diagram of the PASIGLA system. The system is primarily used by two roles: the Health Worker and the Admin. Health workers are responsible for inputting disease case data and medical stock information. Based on this input, the system generates reports that include current and projected trends in disease cases and inventory levels, as well as the risk classification of barangays. On the other hand, the admin handles system configuration and backend management, including user accounts, prediction models, and database maintenance. In return, the system provides the admin with reports related to system performance and data integrity.

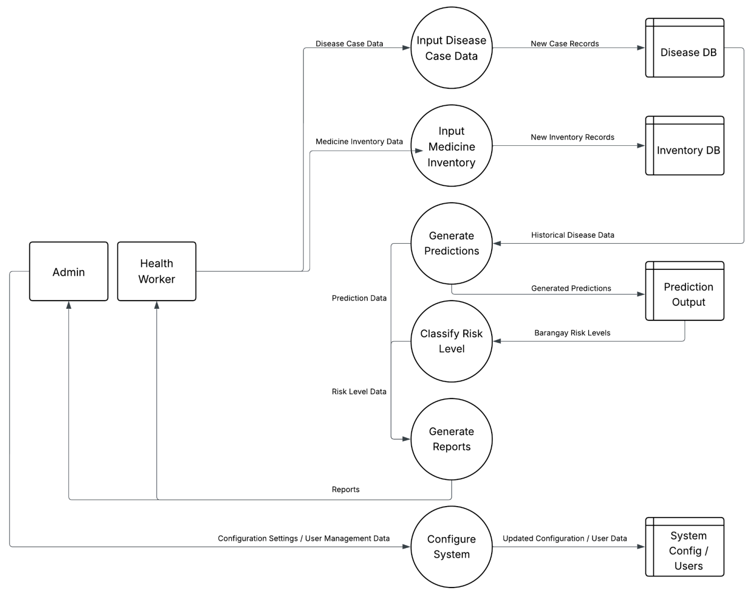

Figure 4: Level 1 Data Flow Diagram

Figure 4 shows the deeper relation of the user and each of the functions of PASIGLA presented by a Level 1 data flow diagram. Detailing operations and functions on a specific and deeper level.

The figure 5 is the Entity Relationship Diagram (ERD) represents the logical structure of the database used in the PASIGLA system. It outlines the core entities, their attributes, and the relationships among them. This model ensures data consistency and supports key functionalities such as disease case tracking, inventory management, prediction generation, and report handling.

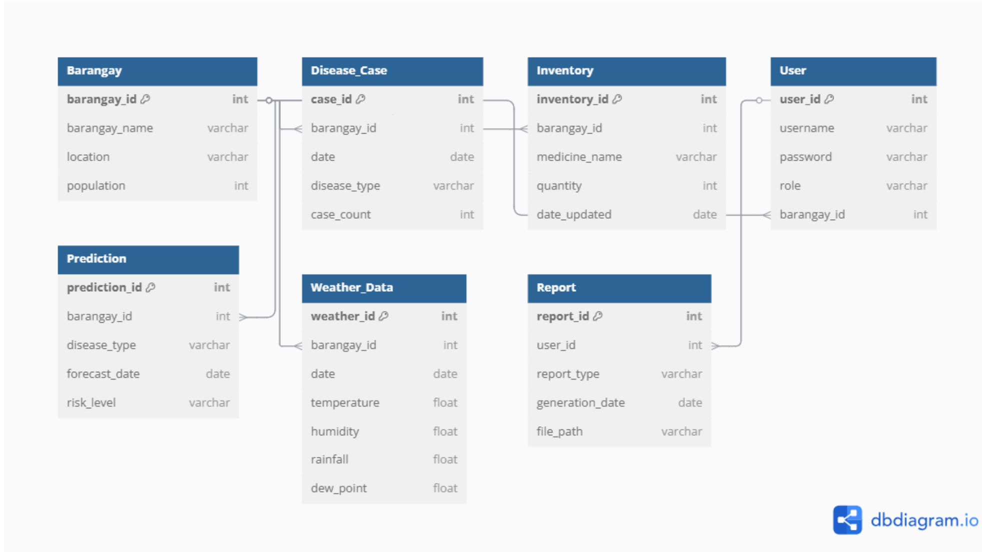

Figure 5: Entity Relationship Diagram

The first table is the Barangay table which serves as the primary or center reference for all other tables. Each barangay is identified with a barangay_id and contains attributes such as name, location, and population. Disease_Case table records reported cases of illnesses per barangay and date. It includes disease_type and case_count which track and analyze historical trends.Inventory table manages the medical stocks. Each record is connected to a barangay and includes medicine_name, quantity, and date_updated keeping track of the latest stock numbers. With a role based system, the user table serves as a storage for user information which includes usernames, passwords, roles, and their associated barangay. The Prediction table stores the predicted output by the machine learning models. Includes disease_type, forecast_date, and risk_level. The Report table stores metadata about the generated reports by the system tracking the report type, creation date, and file path. Lastly, machine learning models require additional information to accurately predict, the Weather_date table is created. This table serves as a storage for an affecting factor for the prediction, common diseases are affected by weather conditions.

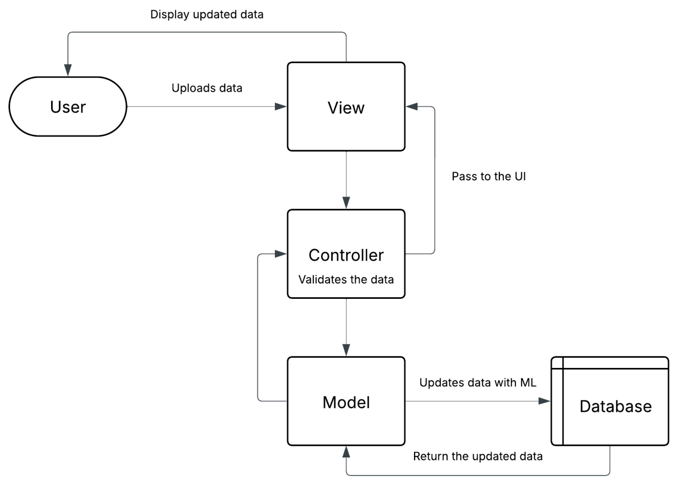

Figure 6: MVC Architecture of Prediction function

The Model-View-Controller (MVC) architecture separates the functionalities into three sections namely the Model, View, and Controller which organize the codebase and improve maintainability.. Figure 6 illustrates the role of each section on the system. The user uploads data, a disease cases count or inventory count, to the view through an input form, the view passes it to the controller to be validated. Once validated, the Controller communicates with the Model, which processes the data and interacts with the database. After the Model updates the database, it retrieves the latest data and sends it back to the Controller. The Controller then updates the View, which displays the processed or predicted results to the user in an understandable format.

2.2.3 Tool to test for Accuracy

The study's machine learning models were divided into two categories: regression/time series and classification. Under the category of regression, where it identifies trends and forecasts upcoming trends, are SARIMA and LSTM. Under the classification categories, where it identifies the risk level of barangay, are Logistic Regression, Naive Bayes, and SVM. Random Forest, XGBoost, and K-NN were added since they can do both functions of the two categories.

Time series models are evaluated through their accuracy of aligning to the available data. Testing done in this study where the data is divided into two sets of historical and current. To achieve a high accuracy rate, the model must align itself to the current data by analyzing and utilizing the historical data given. Evaluation metrics used in this study for this category are Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Average Percentage Error (MAPE), and R^2 score.

- The mean absolute error, or MAE, determines the average difference between the computed and actual values. Additionally, it computes observational error using the same scale that is used to forecast machine learning model accuracy (GeeksforGeeks, 2025).
- Root Mean Square Error (RMSE) is regarded as a superior general-purpose error metric because RMSE is scale-dependent. It is only useful for comparing the prediction errors of various model configurations

for a single variable. It is a metric used to assess how well a regression line fits the data (GeeksforGeeks, 2024).

- The Mean Average Percentage Error (MAPE) allows you to evaluate forecast accuracy across time-series models and uses absolute numbers to prevent positive and negative mistakes from canceling each other out (*IPM Insights Metrics Include MAPE (Mean Absolute Percentage Error)*, 2025).
- R^2 score shows how much of a dependent variable's fluctuation can be accounted for by an independent variable in a regression model (Fernando, 2024).

The machine learning models under the category of regression are evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Average Percentage Error, and R^2 score for regression models were used to train and assess each algorithm.

Algorithm	**MAE**	**RMSE**	**MAPE**	**R^2 Score**
SARIMA	145.11	172.43	81.83%	-5.8265
LSTM	87.23	109.58	57.43%	-1.7570
Random Forest	108.49	131.94	75.92%	-2.9968

XGBoost	78.43	59.51	41.83%	-0.4123
K-NN	68.51	61.33	42.29%	-0.0778

Table 1: Regression Models Evaluation Results

Results indicate that the K-NN algorithm performs better than the other machine learning models with -0.0778 of R^2 score. However, a negative R^2 score dictates poor performance. All machine learning models have negative R^2 score indicating poor performance in interpreting the dataset. With further testing and alteration of the algorithm itself or the dataset. Finding the best model to accurately predict disease cases is important.

These results reflect the entirety of the sample dataset. Due to vastness and difference of cases per barangay, the trend became chaotic and difficult to be analyzed and understood by the regression models, hence resulting in poor performance. However, scaling down the prediction to each barangay, the results of evaluation seems to improve.

Model	**Barangay**	**RMSE**	**MAE**	**MAPE**	**R2**
KNN	Bagong Katipunan	0.384594	0.232346	17.66	-0.18
LSTM	Bagong Katipunan	0.357696	0.237770	21.43	-0.02

Random Forest	Oranbo	0.314945	0.191922	17.05	0.03
XGBoost	Bagong Katipunan	0.510405	0.277458	24.30	-1.08

Table 2: Model-Barangay Evaluation Results

Results indicate that the model performs significantly better if the scope is per barangay. Reason being that the data per barangay is more clean compared to the entirety of the dataset which is chaotic. This led the researchers to a run testing each of the models to each barangay to identify which models predict accurately adapting to the data of the barangay. Presented below are the models most likely suitable for each barangay's historical data:

	Barangay	Model	RMSE	MAE	MAPE	R2
0	BAGONG ILOG	LSTM	3.256760	2.205930	53.217236	0.111369
1	BAGONG KATIPUNAN	KNN	0.384594	0.262346	17.655859	-0.183299
2	BAMBANG	RandomForest	2.086933	1.748272	45.679608	0.036618
3	BUTING	XGBoost	1.312501	0.821548	39.092186	-0.221985
4	CANIOGAN	LSTM	3.748160	2.647644	39.199600	-0.743977
5	DELA PAZ	KNN	3.013697	2.122685	51.432981	-0.678897
6	KALAWAAN	KNN	5.525876	4.131944	95.756173	-0.329227
7	KAPASIGAN	KNN	1.481276	1.092222	39.403704	-0.171675
8	KAPITOLYO	KNN	0.784855	0.643519	27.361111	-0.277844
9	MALINAO	LSTM	0.753390	0.604282	30.216543	-0.201969
10	MANGGAHAN	RandomForest	9.740065	7.944700	36.493012	0.052101
11	MAYBUNGA	RandomForest	3.007378	2.123106	25.091071	0.655363
12	ORANBO	RandomForest	0.290351	0.166178	15.125722	0.172291
13	PALATIW	KNN	3.229973	2.547593	38.660165	0.067466
14	PINAGBUHATAN	KNN	18.062318	14.702778	59.946381	-0.359049
15	PINEDA	KNN	1.069582	0.835648	40.929784	0.530330
16	ROSARIO	LSTM	7.575993	6.697702	40.330077	-0.230640
17	SAGAD	RandomForest	0.841426	0.669583	62.020833	-2.185993
18	SAN ANTONIO	KNN	1.449776	1.083333	39.027778	-1.346253
19	SAN JOAQUIN	KNN	4.220020	3.655000	69.696154	-0.006304
20	SAN MIGUEL	LSTM	3.916759	3.198328	95.013950	-0.305617
21	SAN NICOLAS	RandomForest	0.261576	0.218048	20.357413	0.433748
22	SANTA CRUZ	XGBoost	0.533965	0.442384	33.727273	-0.368568
23	SANTA LUCIA	KNN	6.906180	4.634815	35.672571	0.216236
24	SANTO TOMAS	LSTM	2.160189	1.673517	72.187593	-0.075573
25	SANTOLAN	LSTM	5.598047	4.332335	35.666276	0.371404
26	SUMILANG	KNN	0.841601	0.576420	29.494200	0.220648
27	UGONG	KNN	1.471174	1.208333	40.625000	-0.900407

Figure 7: Best Model per Barangay

As for the classification algorithms, evaluation metrics used are confusion matrix and classification report underlying the accuracy score.

Algorithm	**Accuracy**	**Precision**	**Recall**	**F1-Score**
Logistic Regression	0.67	0.69	0.67	0.66
Naive Bayes	0.61	0.61	0.61	0.60
SVM	0.72	0.73	0.72	0.72
Random Forest	0.72	0.73	0.72	0.72
XGBoost	0.74	0.76	0.74	0.73
K-NN	0.71	0.72	0.71	0.71

Table 3: Classification Models Evaluation Results

Results show that the XGBoost algorithm is the most suitable for classifying whether a barangay is high-risk or low-risk. Accurate classification is crucial in this study, as it serves as the basis for planning and resource allocation—enabling local government units to prioritize interventions in barangays identified as high-risk.

With regards to the inventory management, the system tested two machine learning models, resulting to the following performance evaluation:

Algorithm	MAE	RMSE	MAPE	R^2 Score
SARIMA	721226.53	732255.73	100.00%	-32.4481
LSTM	87.23	109.58	57.43%	-1.7570

Table 4: Inventory Management ML Models Evaluation Results

The results indicate that the tested machine learning models were unfit for the system as seen by their evaluation scores. While both models have negative R-squared scores which they perform worse than simply predicting the mean. LSTM is much lower MAE, RMSE, and MAPE suggest it captures short-term dependencies far better than SARIMA, which fails to generalize due to extreme error spikes and possibly unmodeled seasonality or trend shifts. This indicates that deep learning models like LSTM are more appropriate for inventory items with erratic or surge-driven procurement behavior, while traditional time series models struggle with such volatility.

Evaluation metrics used for classification category are Accuracy, Precision, Recall, and F1-score.

- Accuracy is the ratio of correctly predicted instances to the total number of instances in the dataset (Bonnet, 2025b).
- Precision calculates the proportion of true positive predictions among all positive predictions made by the model (Bonnet, 2025c).

- Recall calculates the proportion of actual positive cases correctly identified by the model (Bonnet, 2025d).
- F1-score balances precision and recall by taking the harmonic mean of the two. When both false positives and false negatives are important, the F1-score is especially helpful (Netguru, n.d.).

2.3 Population and Sampling

PASIGLA, an innovative healthcare platform being developed, targets local health institutions in Pasig City. Software heavily emphasizes machine learning-based illness prediction and pharmaceutical inventory optimization with a strong focus on predictive analytics nowadays. Disease occurrence likelihood predictions at barangay level are crafted from publicly available health data mainly for health professionals. System aims at enhancing accuracy and efficiency of drug inventory management thereby preventing shortages and overstocking in various local health centers simultaneously.

This study will specifically focus on common diseases in Pasig City, grouped into communicable and non-communicable categories, selected based on their prevalence and the availability of relevant data. The targeted illnesses include:

- **Communicable Diseases:** Dengue, Tuberculosis, Influenza-like Illness (ILI), and Leptospirosis;
- **Non-Communicable Diseases:** Diabetes, Hypertension, Cardiovascular Diseases, and Chronic Obstructive Pulmonary Disease (COPD).

Historical health data from 2014 through 2024 will be analyzed heavily using Python-based machine learning algorithms with focus on COVID-19 pandemic years 2020–2021. Algorithms will anticipate disease trends rather quickly and support allocation of medical supplies nationwide with fairly accurate forecasting. Inventory monitoring technologies facilitate real-time pharmaceutical requirement forecasting based on health patterns detected through advanced analytical tools. System development will prioritize usability for health personnel and admin staff alike with slick remarkably user-friendly interface.

Scope of study will be limited to above mentioned diseases and barangays of Pasig City with sufficiently reliable data available for rigorous analysis. Predictions won't be made for individual patients largely upholding stringent confidentiality and privacy standards vigilantly nationwide every single time. System will ignore diseases beyond specified categories and exclude barangays with sketchy health records potentially compromising predictive model accuracy.

Additionally, validity and reliability of system functionality rely on quality and consistency of datasets utilized frequently in analysis processes. Significant data gaps or biases may lead quickly to wildly incorrect forecasts and underestimation of grave potential health risks down the line. Successful deployment and maintenance of such a platform necessitate robust computing infrastructure and regular software updates alongside considerable technical support imposing potential resource limitations on some local health units.

2.4 Data Collection Method

Development of PASIGLA platform was undertaken using various digital tools and programming libraries supporting data processing and machine learning model creation simultaneously. Python became primary programming language owing largely to extensive libraries and sizable support for various data science related tasks. Pandas and NumPy libraries were utilized heavily for preprocessing large volumes of health data including disease case counts and sundry inventory records. These tools facilitated transformation of raw unstructured health datasets into formats usable for in-depth analysis quickly. Scikit-learn, TensorFlow and Keras were utilized heavily for building algorithms that identify trends in disease outbreaks and forecast medication demand effectively. Machine learning frameworks enabled researchers experimenting with various techniques like classification or time-series forecasting ensuring generation of pretty accurate outputs.

Development and testing proceeded rapidly using Jupyter Notebook. Data used in this study came from highly credible and somewhat official sources like Department of Health and various local government units within Pasig City. Datasets spanning 2014 through 2024 were utilized as core input for disease prediction features and inventory management on the platform. Researchers designed a system that was technically efficient and grounded pretty deeply in real-world public health needs by combining various tools.

Research was carried out through a series of meticulously laid out steps ensuring each development phase was grounded deeply in fairly accurate data.

Identifying and collecting health-related data pertinent locally in Pasig City was initial step involved heavily in research process somehow. Historical logs of medicine stocks and usage rates from 2014 alongside recorded cases of dengue and diabetes were amongst data sources utilized. Particular emphasis landed squarely on years 2020 and 2021 reflecting worrisome patterns during COVID-19 pandemic, a rather critical period in global public health.

Meticulous cleaning and preprocessing ensued afterwards with data acquisition now complete. This step's goal was making datasets ridiculously suitable for machine learning algorithms with considerable tweaking and quite a lot of effort. Data scrubbing tasks including eradication of duplicate records and normalization of assorted formats were executed pretty thoroughly afterwards. Data fed into the system were rendered fairly consistent and pretty meaningful thereby ensuring a robust foundation for further analysis downstream somehow.

Researchers proceeded quickly with development of their model after meticulously preparing relevant datasets. Machine learning algorithms like Random Forest and Logistic Regression were trained on data from specific barangays detecting patterns that signal disease outbreak likelihood. Forecasting models were built predictively using past inventory usage data for specific medicines and future demand simultaneously. Models were rigorously tested on various dataset subsets and evaluated using metrics such as accuracy precision recall and F1-score mostly for comparison purposes.

Trained models were subsequently integrated into the PASIGLA system forming the basis of its two primary functions: disease prediction and inventory management optimized rather effectively. Internal testing was performed pretty thoroughly after integration ensuring system outputs were spot on and staff with minimal tech training could use it fairly smoothly. This elaborate procedure verified system functionality and practicality while aligning with real-life needs of various local healthcare providers pretty effectively.

2.5 Data Analysis Procedures

PASIGLA's success heavily relied on extracting meaningful insights from health-related datasets largely through application of various fancy data analysis techniques (Han, Pei, & Kamber, 2011). Supervised machine learning models were utilized by research team on meticulously cleaned historical health records revealing obscure patterns within. Models aimed at linking past disease trends with various environmental factors and demographic info forecast future outbreak risks at barangay level effectively nationwide (Uddin, Khan, Hossain, & Moni, 2019).

Performance metrics were used quite effectively for evaluating effectiveness of predictive models. F1-score offered a balance between precision and recall while accuracy measured correctness of predictions and recall determined detection of actual disease occurrences well (Saito & Rehmsmeier, 2015). Metrics provided a comprehensive view of model reliability and were utilized alongside various algorithms selecting best-performing ones for each illness type accordingly.

Inventory data were subjected to various time-series analysis techniques besides being used for disease forecasting purposes regularly. Seasonal fluctuations and usage patterns of specific medications became readily apparent allowing team detection with considerable ease and surprising clarity (Chatfield, 2000). Predictive models were trained afterwards estimating future stock needs thereby helping health centers make informed procurement decisions subsequently. PASIGLA aimed at reducing medicine waste from overstocking significantly while dodging perilous shortages of crucial medical supplies historically. Results from data analyses were subsequently integrated into platform functionality somehow quite obviously usable and actionable for healthcare professionals (Dorota-Owczarek, 2024).

2.6 Ethical Considerations

Sensitive health-related data research was conducted with rigor under very strict ethical standards throughout entire investigation processes. Datasets utilized in this study were thoroughly anonymized and lacked personally identifiable information altogether (Philippine Health Research Ethics Board, 2017). Researchers utilized only publicly accessible datasets properly cited and sought formal permission from relevant agencies for using localized data where necessary for academic purposes (Republic Act No. 10173, 2012).

System design focused on barangay level functionality rather than tracking individual patient data very specifically within its operational framework. Decision ensured predictions generated by PASIGLA informing broader public health strategies quietly without overly compromising sensitive personal privacy somehow downstream. Researchers took great care ensuring all algorithmic decisions made by system were transparent and explainable to users always (Goodman & Flaxman, 2017). Feedback from community health workers during pilot stage was handled discreetly with utmost respect largely for fine-tuning entire system rather than critiquing individual performances.

Study ensured PASIGLA platform remained effective and trusted amongst healthcare community by upholding stringent data handling and user engagement protocols rigorously.

REFERENCE

Bonnet, A. (2025a, February 26). *Accuracy vs. Precision vs. Recall in Machine Learning: What is the Difference?* https://encord.com/blog/classification-metrics-accuracy-precision-recall/

Bonnet, A. (2025b, February 26). *Accuracy vs. Precision vs. Recall in Machine Learning: What is the Difference?* https://encord.com/blog/classification-metrics-accuracy-precision-recall/#:~:text=Accuracy%20represents%20the%20ratio%20of%20correctly%20predicted%20instances%20to%20the%20total%20number%20of%20instances%20in%20the%20dataset.

Bonnet, A. (2025c, February 26). *Accuracy vs. Precision vs. Recall in Machine Learning: What is the Difference?* https://encord.com/blog/classification-metrics-accuracy-precision-recall/#:~:text=Precision%2C%20often%20referred%20to%20as%20the%20positive%20predictive%20value%2C%20quantifies%20the%20proportion%20of%20true%20positive%20predictions%20among%20all%20positive%20predictions%20made%20by%20the%20model.

Chatfield, C. (2000). *Time-Series Forecasting*.

Dorota-Owczarek. (2024a, January 14). Revolutionize Your Stock Levels with Machine Learning Inventory Management. *nexocode*. https://nexocode.com/blog/posts/inventory-optimization-machine-learning/

Dorota-Owczarek. (2024b, January 14). Revolutionize Your Stock Levels with Machine Learning Inventory Management. *nexocode*. https://nexocode.com/blog/posts/inventory-optimization-machine-learning/

Fernando, J. (2024, November 13). *R-Squared: Definition, Calculation, and interpretation*. Investopedia. https://www.investopedia.com/terms/r/r-squared.asp

Gao, Y., Cai, G., Fang, W., Li, H., Wang, S., Chen, L., Yu, Y., Liu, D., Xu, S., Cui, P., Zeng, S., Feng, X., Yu, R., Wang, Y., Yuan, Y., Jiao, X., Chi, J., Liu, J., Li, R., . . . Gao, Q. (2020). Machine learning based early warning system enables accurate mortality risk prediction for COVID-19. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-18684-2

GeeksforGeeks. (2024, October 11). *RootMeanSquare error in R programming*. GeeksforGeeks. https://www.geeksforgeeks.org/r-language/root-mean-square-error-in-r-programming/

GeeksforGeeks. (2025, May 27). *How to calculate mean absolute error in Python?* GeeksforGeeks. https://www.geeksforgeeks.org/how-to-calculate-mean-absolute-error-in-python/

Gonzalez, M. (2023, September 8). Poor planning and logistics behind P7.4-B wasted, undistributed medicine at DOH. *RAPPLER*. https://www.rappler.com/philippines/department-health-poor-planning-distribution-medicine-wastage-audit-report-2022/

Han, J., Kamber, M., & Pei, J. (2011). Data Mining : Concepts and Techniques 3rd edition Ed. 3. In *Elsevier eBooks*. https://www.scholartext.com/book/88809627?_locale=fr

Hayes, A. (2024, June 27). *Inventory Management: Definition, How It Works, Methods & Examples*. Investopedia. https://www.investopedia.com/terms/i/inventory-management.asp

IPM Insights metrics include MAPE (Mean Absolute Percentage Error). (2025, June 6). Oracle Help Center. https://docs.oracle.com/en/cloud/saas/tax-reporting-cloud/ustrc/insights_metrics_MAPE.html

National Ethical Guidelines for Health and Health-related Research, 2017. (2018).

Netguru. (n.d.). *F1 score: Artificial intelligence explained*. https://www.netguru.com/glossary/f1-score#:~:text=The%20F1%20Score%2C%20named,false%20negatives%20are%20crucial.

Onay, A., & Onay, M. (2019). A drug decision support system for developing a successful drug candidate using machine learning techniques. *Current Computer - Aided Drug Design*, *16*(4), 407–419. https://doi.org/10.2174/1573409915666190716143601

Philippine Health Research Ethics Board, Reyes, M. V. T., Alora, R. A. T., De Castro, L. D., Jimenez, E. B., Manalastas, R. M., Santos, E. O., Tomas, C. V., & De Castro, L. D. (2017). *NATIONAL ETHICAL GUIDELINES FOR HEALTH AND HEALTH-RELATED RESEARCH 2017*. Department of

Science and Technology - Philippine Council for Health Research and Development. https://www.pchrd.dost.gov.ph/wp-content/uploads/2022/03/Annex-5.-National-Ethical-Guidelines-for-Health-and-Health-Related-Research-2017-1.pdf

Republic Act No. 10173: Data Privacy Act of 2012. (2012, August 15). *Official Gazette of the Republic of the Philippines*. https://www.officialgazette.gov.ph/2012/08/15/republic-act-no-10173/

Saito, T., & Rehmsmeier, M. (2015). The Precision-Recall Plot Is More Informative than the ROC Plot When Evaluating Binary Classifiers on Imbalanced Datasets. *PLoS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019a). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, *19*(1). https://doi.org/10.1186/s12911-019-1004-8

Uddin, S., Khan, A., Hossain, M. E., & Moni, M. A. (2019b). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, *19*(1). https://doi.org/10.1186/s12911-019-1004-8

Venkatesh, M., & Raju, B. (2023). MULTIPLE DISEASE PREDICTION USING MACHINE LEARNING, DEEP LEARNING AND STREAM-LIT.

International Research Journal of Modernization in Engineering Technology and Science. https://doi.org/10.56726/irjmets42818

Zwaida, T. A., Pham, C., & Beauregard, Y. (2021). Optimization of inventory management to prevent drug shortages in the hospital supply chain. *Applied Sciences*, *11*(6), 2726. https://doi.org/10.3390/app11062726