SSRG International Journal of Electronics and Communication Engineering
ISSN: 2348-8549/ https://doi.org/10.14445/23488549/IJECE-V12I1P107

Volume 12 Issue 1, 92-103, January 2025

Original Article

Enhanced Data Analysis from GIS as a Smart City by Machine Learning

Ahmed Anwer Al Jumaili[1], Kifah Tout[2], Zaid F. Makki[3]

[1,2]Lebanese University Doctoral School of Science and Technology, Beirut, Lebanon,
[3]Alshaab University, Iraq.

[1]Corresponding Author : ahmed.aljumaily @ ul.edu.lb

Received: 11 November 2024 Revised: 17 December 2024 Accepted: 08 January 2025 Published: 30 January 2025

Abstract *- Smart cities have recently expanded and become a phenomenon sought by urbanized societies. This expansion increases the need for applications to manage these systems efficiently. In this study, we present an approach that integrates Geographic Information Systems (GIS) with one of the Machine Learning (ML) algorithms in order to enhance the analysis of important data that helps in developing smart cities, such as traffic, environmental monitoring, and resource allocation for making important decisions. The well-known classifier Support Vector Machines (SVM) help classify classes and recognize patterns, especially when adding weights that affect the features extracted from the data in the standard dataset. Due to the integration of Artificial Intelligence (AI) techniques with GIS, planning smart city infrastructure and predicting future trends in forecasting improved. Urban management in smart cities is more dynamic through the proposed approach. The study proved the worthiness of the proposed method through good results, as the prediction accuracy reached 90% and high results for the rest of the evaluation criteria. This research paves the way for taking advantage of artificial intelligence techniques by integrating them with GIS.*

Keywords *- Smart city, Support Vector Machines (SVM), Geographic Information Systems (GIS), Machine Learning (ML), Prediction.*

1. Introduction

The development of technology and the rise of smart cities came in response to the need for sustainable urban development, the ability to manage resources efficiently and improve quality of life. There are many sources of data that smart cities benefit from, including widespread sensors, Internet of Things devices and all the data generated by people for distinguished decisions that help urban life [1].

Geographic Information Systems (GIS) play an effective role in smart cities by correctly analyzing this information and simulating the preparation of maps. These technologies are essential for creating urban planning and managing traffic and the environment in general [2].

Modern cities are complex systems interspersed with engines to accelerate the development of human society. These cities consume a lot of natural resources and produce more waste than other lifestyles. This leads to changes in the global climate and undesirable environmental changes. According to studies, in 2050, the urban environment population is expected to reach 67% of the world's population, or an estimated 6.3 billion people. This leads to increased rates of disasters and risks to the population [3].

Data sources in smart cities depend on a variety of different sources. Data comes from different sources such as sensors, where different sensors such as surveillance cameras, motion, light and sound sensors, and Internet of Things devices are used to collect accurate and immediate data about smart cities. Analyzing this data is a source of traffic and congestion information, as well as weather conditions, air and water quality and pollution, noise levels and other important measurements for smart cities [4].

The second source of data is government data, available open data and data from private institutions, where data is collected from various government platforms, databases and sources, especially municipalities, ministries and government agencies. These sources include a huge set of data on the population and information related to the government and private services provided to them, as well as health and education services, smart city infrastructure maps, geographic data, economic information, population census and urban statistics [5]. The third source of data in smart cities is satellites, which provide live data on traffic in smart cities and monitor energy consumption to improve efficiency. Satellites also provide data on land available for sustainable urban development and current land uses. Data sources include

Geographic Information Systems (GIS) that provide detailed spatial insights into urban planning and land use [6].

The fourth source of data is social media, which is a source of big data, as well as smartphones and mobile devices, where data collection related to public interactions on social media platforms is used. This helps in understanding the needs and interests of residents, communicating with them, and knowing their opinions and comments about smart cities and their services [7].

Although data is obtained from GIS, this data is very large and complex, which poses a great challenge to analyze using traditional methods. Hence, Machine Learning (ML) provides a way to automatically learn this large data and rely on prediction to know the future course of data and solve all the challenges here [8]. Machine learning can analyze large data obtained from GIS and reveal hidden patterns that help improve the functions of the smart city, enabling it to manage more efficiently and adapt to immediate responses.

Of course, after collecting this data, it is necessary to analyze it and transform it into valuable information for smart cities. This goal requires using machine learning and artificial intelligence techniques to extract value from this data to achieve the goals of smart cities in improving urban life, raising the quality of life, and providing better services to residents and visitors of these cities. Therefore, developing smart city infrastructure is essential to make the most of collecting and transferring data from its sources.

This manuscript discusses using one of the artificial intelligence techniques represented by machine learning to improve the big data coming from the GIS system in smart cities, taking into account applications such as energy management, public services, and transportation management. The aim of this study is to process big data through machine learning and improve the outputs to serve the automation of applications in smart cities and facilitate services. One of this study's main challenges and limitations is the integration of machine learning and geographic information systems to achieve the best services in smart cities.

2. GIS in Smart City

Over the past years, GIS has evolved from a traditional science of managing geographic maps to a fast-paced technology that is a source of the most important information that has contributed to the sustainability of smart cities. The GPS system works to obtain data from various sectors and stores and analyzes it with the help of other technologies and applications to improve smart city data [9].

There are many elements that need to be achieved to create smart cities, and an integrated approach must be taken to accommodate these elements, including green energy, smart construction, livable housing, sustainable transportation, water resources, risk reduction and more. One of the main elements that this study focused on is GIS, which is the largest source of information and is increasing dramatically, as in the examples of smart cities (Amsterdam and Dubai). Applications from giant companies (IBM, Google, and Apple), such as software and maps, have helped attract users to urban cities, such as smart sensors. Modern technology has helped the rapid development of smart cities, which has been the basis for attracting developers and hobbyists and absorbing their skills in the field of building smart cities [10].

GIS is the most important source of information gathering and a hub for integrating data from all aspects, such as knowing the sources of problems, the context surrounding its possibilities, what solutions should be designed, and what scenarios can be followed in a dynamic environment. Figure 1 represents geospatial knowledge and its development with the establishment of smart cities and their complexities.

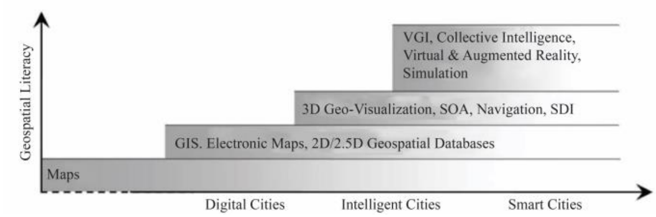

Fig. 1 GIS and related to smart city development [11]

GIS is a technology for data acquisition, analysis and geographic imaging. Through smart cities, these systems help improve the following applications:
Urban planning, which contributes to the design of cities and lands through maps to develop planning and decision-making for housing

Transportation and traffic management, which tracks the movement flow of transportation networks to predict appropriate routes and reduce traffic congestion.
Environmental monitoring regarding water and air, analysis of climate impacts on the urban environment and prediction of potential disasters.

Rapid emergency responses to control waste and allocate population resources.

This requires a method that outperforms traditional methods in processing data in real-time, and here, the need arose to use artificial intelligence technologies and their applications, such as machine learning, to develop smart cities.

2.1. Machine Learning in GIS

Machine learning (ML) is one of the most important applications of AI that we will take into consideration in this study. To review the applications of ML in GIS, it is first necessary to understand the basic concepts related to this field. Below, we present some basic principles and definitions. Artificial intelligence techniques are used in many applications such as military, industrial, security, etc. [12,13,14]. AI techniques have helped in analyzing data obtained from GIS, which is often necessary and has a large security and economic dimension.

ML is an application of artificial intelligence that enables systems to learn independently and improve their performance through experience and expertise without the need for specific programming. This field focuses on developing computer programs that can access and use data in learning [15].

This process begins with observations or data, such as practical examples, direct experiences, or instructions, where this data is used to discover patterns and make more effective decisions based on the examples provided. The main goal of machine learning is to enable machines to learn independently without human intervention and to adapt their actions according to the results they reach.

Machine learning algorithms are usually classified into supervised and unsupervised algorithms. However, this classification is very broad and does not cover all available methods [16].

Supervised ML algorithms can use knowledge gained from past experiences and training on pre-defined examples to predict future events using unseen data. These algorithms start by analyzing a set of training data (pre-defined examples) to predict possible output values. After sufficient training, the system is able to provide appropriate predictions for each new input. In addition, the algorithm can compare its results to the correct outputs, identifying errors to modify and improve the model. Examples of such algorithms include Support Vector Machine (SVM), Decision Tree, Random Forest, KNN algorithm, and Regression [17].

Unsupervised ML algorithms are used in cases where the training data is unlabeled or unlabeled. These algorithms aim to understand how the system can infer a function that describes the hidden pattern in the unlabeled data. Although these algorithms may not determine specific results, they explore the data and can infer and describe hidden structures in the unlabeled data. Examples of such algorithms include Apriori, K-means, and Expectation-Maximization (EM) [18]. Semi-supervised ML algorithms fall somewhere between supervised and unsupervised algorithms and are trained on both labeled and unlabeled data.

Typically, a small portion of the data is labeled, while a larger portion is unlabeled. Systems using these algorithms can achieve a high level of accuracy. Semi-supervised learning is preferred when labeled data requires specialized and efficient resources to obtain, as producing such data is expensive and time-consuming. In contrast, accessing unlabeled data typically does not require additional resources [19].

Machine learning is considered a part of artificial intelligence, which is considered a main title, and in turn, it consists of a smaller part that contains it, which is deep learning, as in Figure 2. For each application, artificial intelligence algorithms work to solve it, and one of them can work efficiently, and another may not work with the same efficiency.

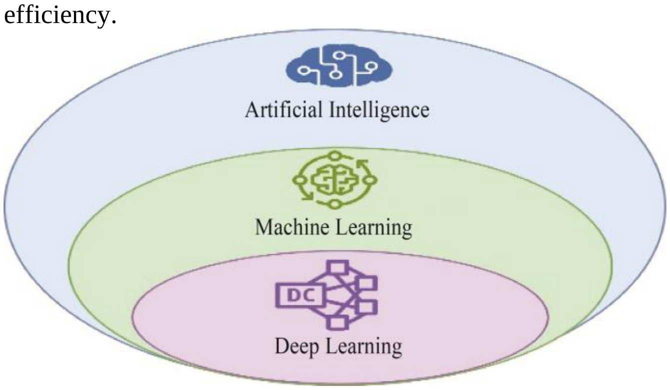

Fig. 2 Relation among AI, ML and DL

ML allows for the analysis of large amounts of data and typically provides faster and more accurate results for identifying profitable opportunities or high risks. However, machine learning can require additional time and resources to ensure it is properly trained.

Machine learning relies on curated data that is used to analyze and build the learning model, which means there is a need for a suitable set of data that can be used effectively in the learning process.

Many ML algorithms are suggested in the literature, and choosing the most suitable algorithm for a particular problem depends on a set of characteristics such as speed, accuracy, training time, prediction time, amount of data required for training, type of data, ease of implementation, etc. Often, the time factor is of utmost importance, especially in GIS applications, as shown in Table 1 [20].

Table 1. Time complexity of some ML algorithms

Algorithm	**Learning**	**Predicting**
Regression	O(p2n+p3)	O(p)
Decision Tree	O(n2p)	O(p)
Random Forest	O(n2pnt)	O(pnt)
Naïve Bayes	O(np)	O(p)
SVM	O(n2p+n3)	O(pnsv)
KNN	---	O(np)
K-means	O(npk+1)	O(k)

To avoid dependency on certain conditions, analyze algorithm runtime for asymptotic sense. Thus, n represents the training number, p is the feature number, nt is the tree number, nsv is the support vector number, and k represents the cluster number. The complexity of ML is calculated according to the table.

Learning time is the time required to train the model using the dataset, and it depends on the size of the data and the type of algorithm used.

Prediction time is the time required to test the model using a new dataset or predict unseen data, and also varies depending on the size of the data and the type of algorithm used.

In most cases, about 80% of the dataset is allocated for training, while the rest is used for fine-tuning and testing. It is worth noting that the training phase is often performed offline, which makes prediction time even more important for developers.

In general, the above criteria can be used to select a number of suitable algorithms, but it is difficult to determine the best algorithm at the beginning. Therefore, it is preferable to follow an iterative approach to work. A set of potential algorithms can be selected from among the machine learning algorithms and tested on the data by running them in parallel or serially. Then, their performance is evaluated to select the most effective algorithm.

In the following section, the proposed method will explain the reason for using the SVM classifier and how to enter GIS information into the classifier to analyze it and find the best ways to achieve a smart city.

2.2. Applications of ML in GIS of Smart City

In smart cities, machine learning works by obtaining data from GIS and analyzing it to improve many applications, including [21]:

- Traffic and transportation systems process traffic flows, movements, road congestion and vehicles in real-time to improve their management using artificial intelligence by predicting traffic patterns and traffic timings through famous classifiers such as SVM or convolutional neural networks.
- Energy resource management, which is one of the basics of smart cities, controlling energy resources, using clean energy and ensuring its distribution through predictive patterns and according to peak times and good distribution and waste management, often using a classifier such as SVM to predict such technologies and calculate factors such as climate, preparation times and historical work history.
- Monitoring environmental sustainability to control pollution and determine its spread patterns to predict future risks, this requires data from Internet of Things sensors and uses deep learning to predict the spread of pollution and rely on weather and spatial factors.
- Managing infrastructure monitored by automated learning, such as roads, bridges, and buildings and using GIS to access infrastructure and predict maintenance times and the expected life of infrastructure.
- Maintaining safety and responding to emergencies to predict crime scenes, places, and times of natural disasters. Collecting and analyzing data on population density, the geographical nature of the area and the economic situation gives a clear picture to predict the best actual response.

3. Related Work

Geographic information systems play a major role in obtaining data to improve the design of smart cities. GIS works to prepare a huge amount of data, especially related to roads, transportation and traffic, to enable traffic movement [22]. Traditional methods of analyzing data take a long time to process and take actions and decisions regarding them, and this is not feasible with applications that operate in real time [23]. Integrating GIS technology and machine learning is an emerging field that needs more research and development. Many studies have considered integrating artificial intelligence applications with GIS applications in the design of smart cities [24]. Predictive analytics played a crucial role in making decisions based on improving services related to GIS and building modern smart cities [25].

Higher education works to enhance the decision-making needed by many smart city facilities, which mainly depend on analyzing and processing data very quickly, even if it is complex. GPS technologies and big data have been used to help design several smart cities, including Dubai, which uses several artificial intelligence technologies to design traffic services and improve health services [26]. Machine learning technologies such as SVM were used to classify data and make decisions that contributed to reducing the effects of natural disasters in several cities in India and had a major impact on increasing the migration of people to urban areas enjoyed by areas such as Delhi [27]. An applied study combined machine learning technologies and big data from GPS systems that processed that data and were the reason for establishing urban cities in record time due to processing economic data and infrastructure and predicting the best investments and the best population cover as in Riyadh, in the Kingdom of Saudi Arabia [28].

The well-known classifier used SVM to classify data collected for several years in Amsterdam, in which GPS had a clear and effective impact and built a smart system to choose the places where the technologies work and attract tourists in an urban way that helped provide resources for people and remove waste in smart ways that do not affect the environment [29]. In a study on the largest cities in the Kingdom of Saudi Arabia, which smart technology helped build, King Abdullah City is considered the largest economic city in the Kingdom, collected data from all sources, including GIS, especially in the subject of energy and security, to improve its urban planning [30]. Environmental sustainability faces major challenges in developing cities, including pollution, waste management and its derivatives. GIS monitors these changes on a wide environmental scale to draw risk areas, but this is not feasible with traditional methods. Many studies have begun applying artificial intelligence and neural network techniques to process big data [31] to assess risks, such as flood prediction in some countries, which have avoided disasters [32]. There are also successful models in Makah and Madinah to serve pilgrims, Umrah performers and visitors to religious sites, where data analysis is used to improve transportation, security, public safety and healthcare services based on smart systems to monitor crowding and jostling, as well as monitoring public safety and emergency health cases, during the performance of religious rituals.

The collected data is also used to make strategic decisions to improve the quality of services each year [33].GIS data has been used in many smart cities, such as Seoul in South Korea [34], which is considered one of the cities that have benefited from the automation of traffic and energy consumption to improve services provided to urban citizens and predict environmental disasters. Many studies have investigated the use of various applications, including artificial intelligence, in designing and improving smart cities. Figure 3 shows the number of studies focusing on smart cities over the past years.

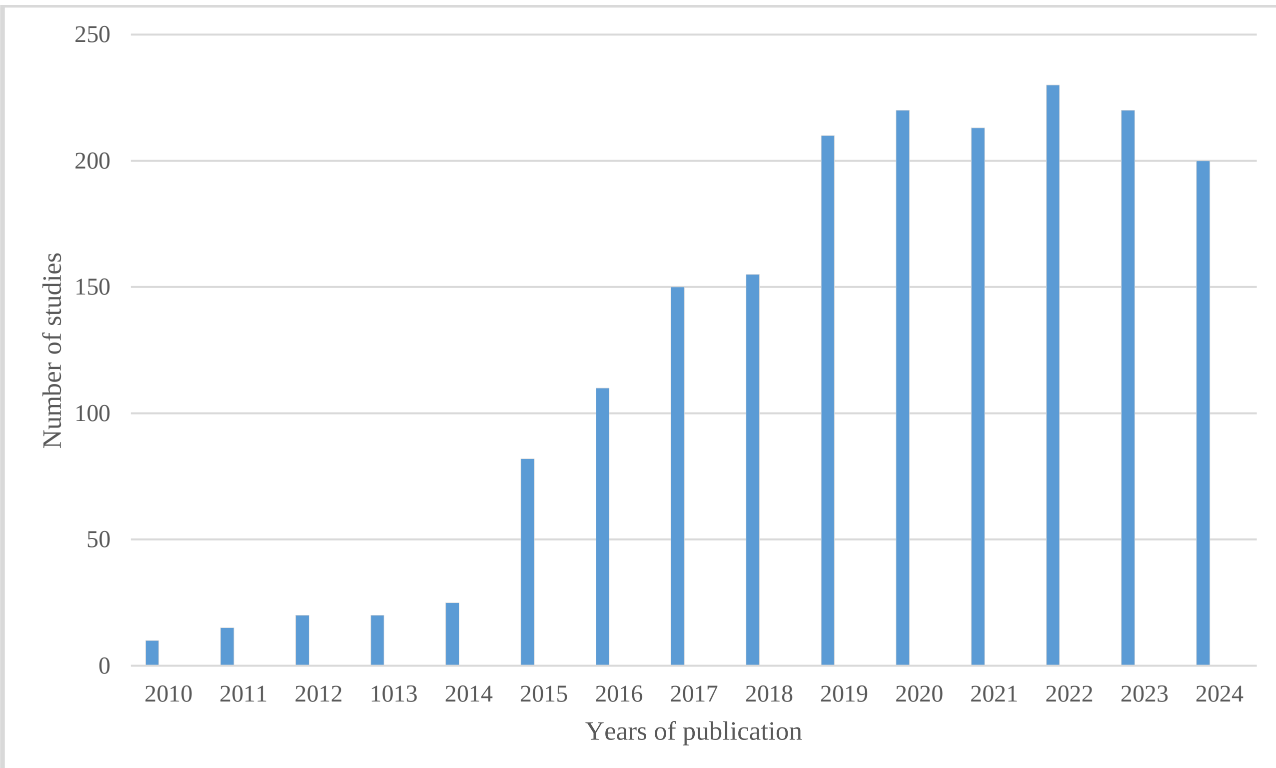

Fig. 3 Previous studies on smart cities and its applications [35]

4. Proposed Method

The process of improving and designing smart cities, according to the proposed methodology, includes integrating one of the most important machine learning techniques, the Support Vector Machine (SVM), with the most famous data acquisition system, GIS. It includes several steps, the most important of which is data acquisition, then pre-processing and training the model to choose the decision in real-time. The proposed methodology can be applied to any application, whether traffic and transportation management, environmental monitoring, land use planning, etc. The general framework is illustrated in Figure 4. The first steps in data collection are historical data recorded in a standard dataset about the land's topography and previously recorded environmental information in addition to spatial and non-spatial information about the city's infrastructure. Data sources here are multiple, including GIS data such as maps and images that come via satellite. All coordinates are in a dataset in addition to information from the Internet of Things. Data is obtained from sensor devices such as traffic cameras and weather stations; the information is instantaneous. Fixed information such as traffic patterns, population density, and terrain is also available. In general, it is of two types: historically fixed and variable in the long term, and information that occurs instantaneously in real time, as shown in Figure 5.

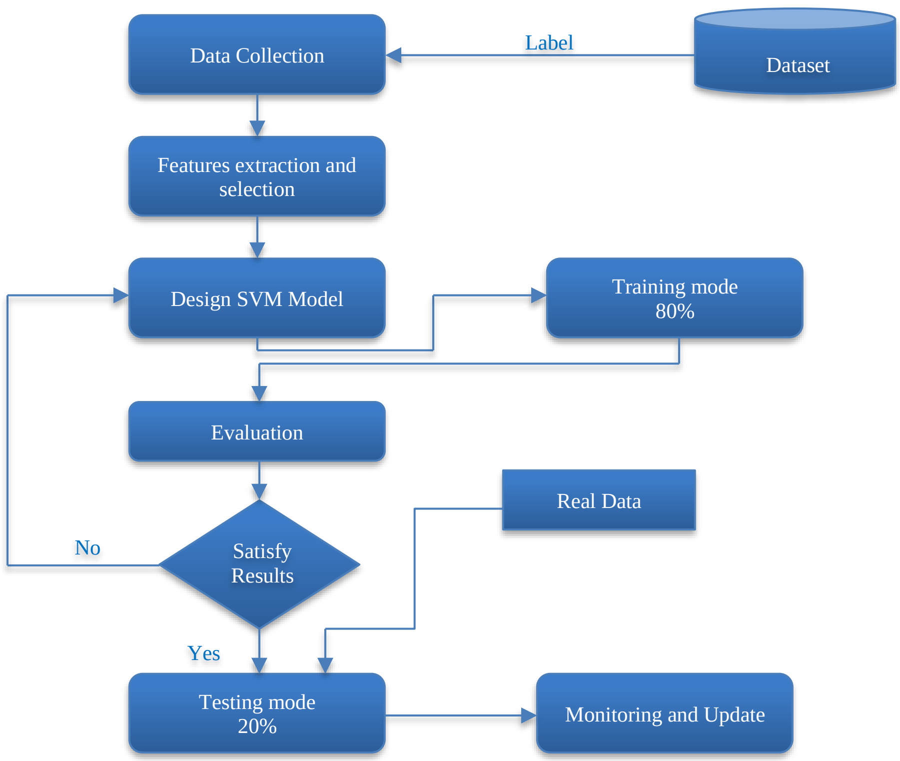

Fig. 4 General framework of the proposed method

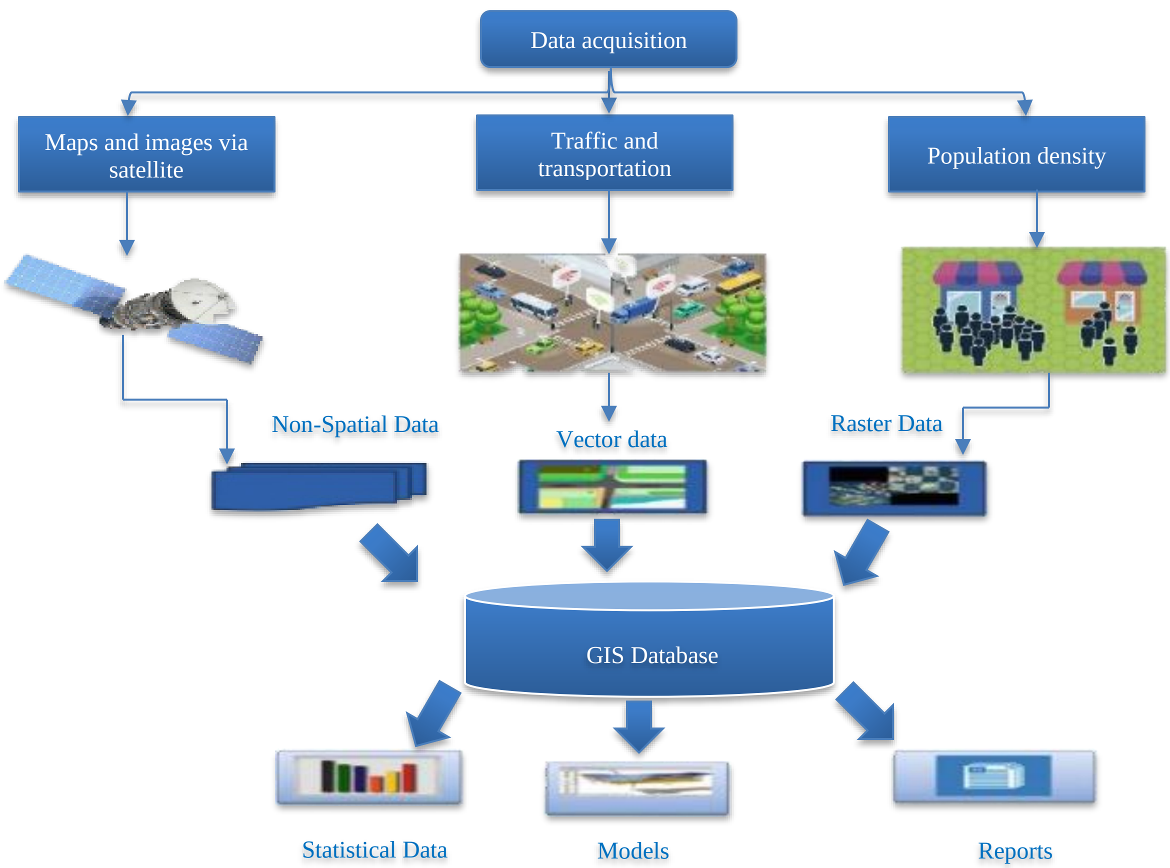

Fig. 5 Type of data acquisition from the dataset

Features extraction and selection represent the data before processing, which are stored in vectors, and these vectors are the only way to enter data into the machine learning algorithm (SVM). Vectors contain features extracted from the data in the dataset and are often a two-way vector so that the classification process can be easily done.

When the vectors containing the features are unified, the process of designing the SVM classifier begins to be suitable for classifying the features in the vectors. The training begins with the data in the database, and the classification is binary; for example, traffic congestion is either high or low (true or false), or sometimes multiclass, as in the land selection classification.

Then, the kernel function is chosen for linear separation between the classes, and here comes the contribution so that the weights help classify the class and the classification is as close as possible to the margin boundaries. The classifier draws a map of the distribution of features in the two classes and which belongs to the other.

There is a set of steps that the classifier takes to determine feature classes. In order to train the classifier, samples are taken from the known classes to determine the unknown ones that do not contain a label. Therefore, when the pre-trained classifier is executed on unknown or new activities, it is known which class it belongs to.

Repeating training leads to increasing the accuracy of the output and thus creating a fully classified class. Classification takes the form of statistical calculations, utilizing the acquired information and determining the features belonging to any class.

First, the classes and their belonging to any group are identified in training, and they are used through the classifier to analyze them by features. As following:

$$D = \{(x_i . y_i) | x_i \in \mathcal{R}.\ y_i \in \{-1.1\}\}_{i=1}^{n} \tag{1}$$

Such as, each xi considers an n-dimensional real feature vector and yi considers either +1 or -1, which refers to the classes by which the feature of xi is included. This helps to determine the best-split line (in terms of linear) or curve (non-linear) that groups the features, including yi = 1 (belong class A) and those of yi = -1 (belong class B). The best separate line between two classes can be found in the case of training data, which is linear behavior (Figure 6).

The weight wi belongs to the feature can be found in the next step. During sorting features in certain vector <w_1.w_2.….bn>, then can calculate the weight cent red in each class by,

$$\hat{C} = \frac{1}{w} \sum w_i x_i y_i \tag{2}$$

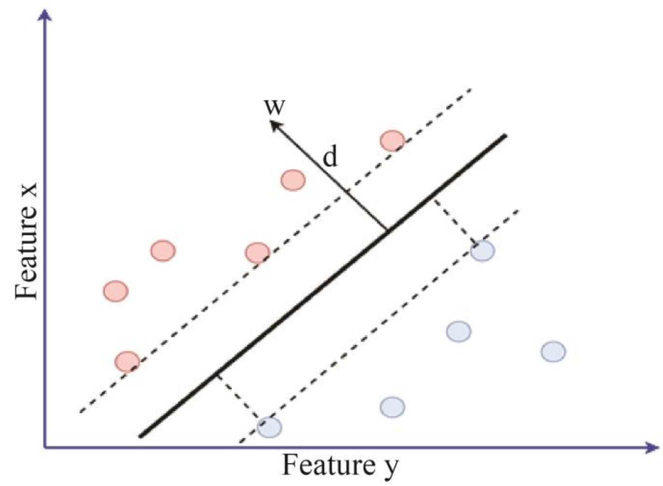

Fig. 6 Linear classifications of selected class

In this step, we can group similar features in order to separate them into specific categories. The process continues for all categories, including non-overlapping categories that can be easily separated by a straight line. It is separated into two classes, A or B. The matching strategy was adopted in non-linear classification, which contains a section of features that overlap between classes such that their place is in a certain class while, in reality, it belongs to another class. The boundaries of the categories can be preserved to predict the best-dividing line to unify the network paths in the distribution of data, as shown in Figure 7.

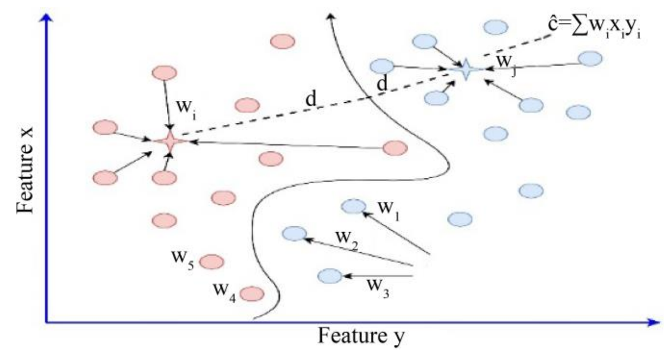

Fig. 7 Non-linear classifications of features

After the algorithm is trained on 80% of the selected samples from the dataset until we reach good percentages of results and expectations according to the label attached to the sample, we test the algorithm without labels on the remaining 20% to evaluate it and know its optimal performance. The prediction accuracy, precision and recall are calculated to evaluate the proposed algorithm based on the weights of SVM and to know the algorithm's ability to classify traffic routes or, predict housing patterns or predict risks, which will be discussed in the next section of the results.

5. Results and Discussion

The practical part of this study is very important, through which the work is evaluated and whether the proposed algorithm can be relied upon. First, we note that the algorithm is trained on real data derived from a standard dataset containing various types of networks, particularly particularly complex ones. We divide the original data into two branches: training and testing (Figure 8). Therefore, the training group is divided into training and validation. The final division of the data is in order to verify the validity of the test data set. The division necessarily has a set of restrictions, and in order for the test data to be good, we do this procedure. When the test data is large, the data must be verified before starting the test, and hence, training depends on the marked data, which often contains all the network features or any path in it.

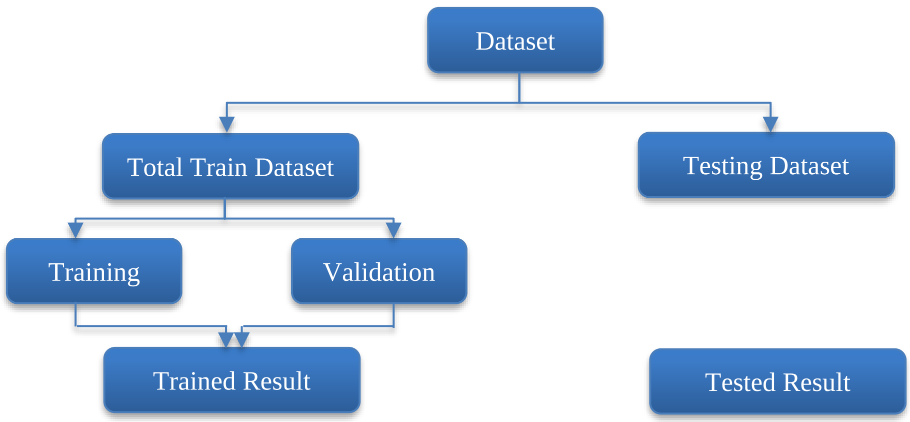

Fig. 8 Architecture of dataset

Features are extracted from GIS information in a given dataset to be classified into vectors for work on them. Each feature has an understanding of the proposed algorithm in order to work on it, as shown below.

```python
def encode_categorical_features(df):
    """
    Checks if columns in the DataFrame are categorical and encodes them numerically.
    Args:
        df (pd.DataFrame): Input DataFrame.
    Returns:
        pd.DataFrame: DataFrame with categorical columns encoded.
    """
    for col in df.columns:
        if pd.api.types.is_categorical_dtype(df[col]):
            df[col] = df[col].cat.codes
        elif pd.api.types.is_object_dtype(df[col]):
            df[col] = df[col].astype('category').cat.codes

    return df
```

The data is collected to be interpreted later, and the collection mechanism represents one of the most important processes that must be adhered to. It works based on data derived from the GIS, such as the history of traffic flow, the topography and terrain, the environmental change, and the population density.

clusters_model= array ([1,0,4,…,0,0,4], dtypeint=32).

In classification, it is generally important to determine the performance evaluation, and it is very useful, especially in the difference in classifications during the training phase. It measures the accuracy of the work of the proposed classifier and the extent of its success. The confusion matrix, or the contingency table, measures the accuracy of the classifier through the columns that represent the predicted and the rows that represent the actual, as shown in Figure 9.

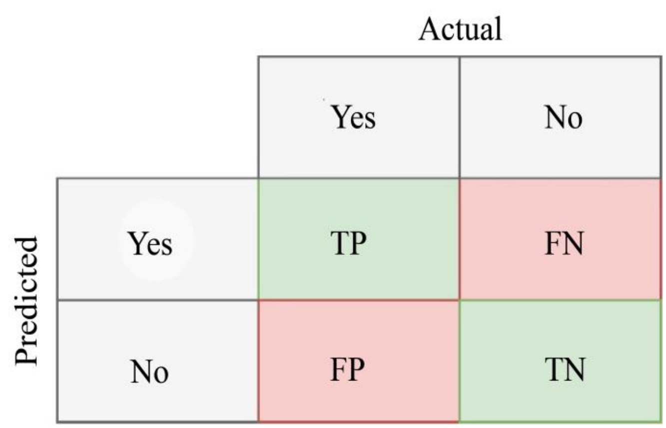

Fig. 9 Confusion Matrix representation where (TP) is TRUE POSITIVE, (TN) is TRUE NEGATIVE, (FP) is FALSE POSITIVE, and (FN) is FALSE NEGATIVE

The accuracy can be found by the Equation:

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \tag{3}$$

Precision tells us what proportion of proper path we detect and have found in the network and can calculated by the Equation:

$$Precition = \frac{TP}{TP+FP} \tag{4}$$

These criteria and others can be illustrated in the following pseudo code:

The results can be translated using the confusion matrix to be clearer and more realistic and to compare to the predicted results, as in Figure 10.

```python
def Report(cm_classes):
  repo = []
  for i in range(len(cm_classes)):
    tp = cm_classes[i,0,0]
    fn = cm_classes[i,0,1]
    fp = cm_classes[i,1,0]
    tn = cm_classes[i,1,1]

    acc = (tp + tn) / (tp + tn + fp + fn) #Accuracy

    precision = tp / (tp + fp) #PPV

    recall = tp / (tp + fn) #Sensitivity

    f1_score = 2 * (precision * recall) / (precision + recall) #F1-Score

    info = {'ACC': acc, 'Precision': precision, 'Recall': recall, 'F1-Score': f1_score}

    repo.append(info)

  return np.asarray(repo)
```

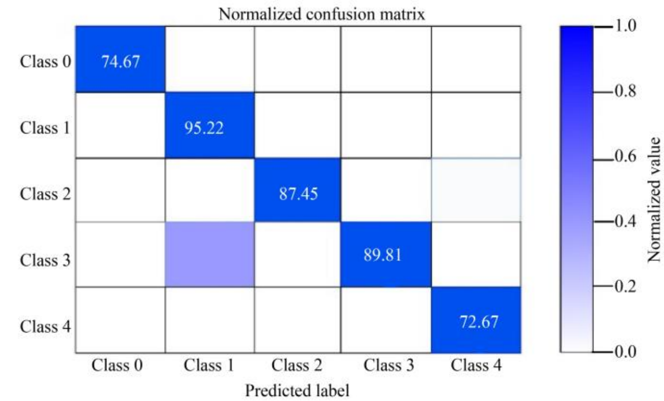

Fig. 10 Confusion matrix of the proposed algorithm

Table 2. Performance of SVM classifier with binary and multiclass

Classification type	Accuracy %	F-measure %	Recall %	Precision %
Binary classification	72.2	61.6	62.8	59.8
Multiclass	81.3	76.9	75.3	73.1
Weighted classification	89.7	78.8	77.4	78.3

The dataset contains 2467800 samples, representing 80% of the dataset during training. And the rest 20% will be in testing mode.

The well-known SVM classifier performed the classification for multiple classes. Classification in the case of the binary class is in the form of a damaged class and a non-damaged class. The results are used to examine the traffic congestion in the affected areas. It is based on the time that can determine the possibility of information reaching the teams that care for the information on external variables, weather and climate. Training is done on the pre-processed data, the changes are saved, and the test is performed on another part of the data. Table 2 shows the training on the 2467800 sample given from the dataset.

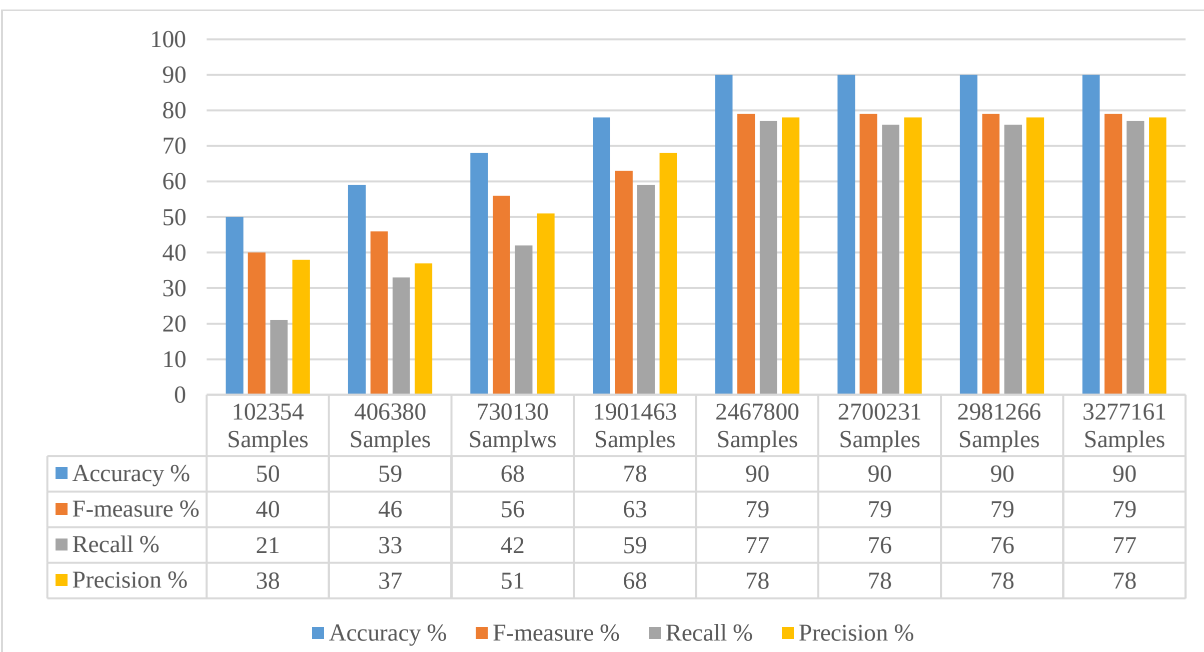

Fig. 11 Evaluation of algorithm within different samples from the dataset in training mode

In the table, results such as precision, recall and average were obtained for data that can be interpreted in multiple ways and for the areas most affected by the information change. Among the expectations are the location of the sensors to be changed according to the new predictions and considering the change in topography and the management of energy sources to be changed according to the predictions we get. The data that was trained on was the most consistent with the real data, so in the test stage, there is a conviction that the results are almost real and can be relied upon. When classifying using the binary class method, the result is weaker and adding weight to the classifier increases the accuracy of the result.

The nationalities varied in the areas where the displaced people were located, including many nationalities and ethnicities for each governorate. The statistics taken by previous researchers in the past years are shown in Figure 10. After applying modern GIS technologies and weight to the SVM, Figure 11 shows the training mode of the system starting with 102354 samples from the dataset until it reaches 3277161 samples.

We see that the specified number of training data is that the more data that is trained on, the more efficient the result becomes until it reaches saturation. Any increase does not serve the system but only increases the time used for training, which is, in fact, very long during the training phase and reaches 12 hours and 34 minutes when executing 2467800 sample data and increases logarithmically after this number. As for the time taken for execution in test mode, it is negligible because the algorithm is pre-trained. The time consumption in training mode can be illustrated in Figure 12, which explains how the training takes a huge amount of time when processing.

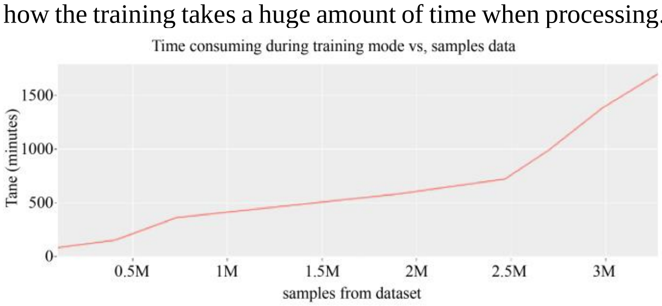

Fig. 12 Execution time of proposed algorithm

6. Conclusion

This study highlights the improvement of data analysis that leads to decision-making in smart cities using GIS and one of the most important machine learning algorithms, which is SVM. Through the big data we obtain from GIS, we address urban challenges using machine learning algorithms for matters such as intelligent traffic management and resource allocation that leads to efficient land planning and environmental monitoring and its choices.

Through the big data that is analyzed using the machine learning algorithm, patterns can be predicted that are of great benefit to improving urban processing. Smart solutions make resources more efficient in smart cities for sustainable urban development. Features are extracted from the big data collected in a standard dataset containing labels to classify them using SVM. Weights are added that effectively impact classifying categories to train the algorithm on the data and evaluate it. The accuracy rate of the proposed method reached 90%, as the proposed method proved its worth through the criteria that the results reached. The proposed methodology provides a scalable model to adapt to smart city data, which leads to more innovations in this field.

Future work

The integration of GIS technology and machine learning algorithms is effective when applied to smart cities and can take an influential direction if it uses real-time data to take data from sensors and IoT devices to process it and make decisions accordingly. SVM technology can be linked to cloud information to control the volume of urban data and expand distributed computing technologies or architecture technology. Using other technologies, such as deep learning, can effectively impact and improve efficiency to increase the ability to predict energy consumption, future traffic routes, or better waste management. The analysis can be more comprehensive, especially if economic, social and demographic data are integrated to enhance urban planning.

Acknowledgement

We would like to express our gratitude to all those who supported this manuscript and our educational institutions (Lebanese University Doctoral School of Science and Technology) for their technical support. We are also grateful to the reviewers for their comprehensive evaluations and constructive suggestions, which in turn improved this manuscript.

References

[1] Emilio Costales, "Identifying Sources of Innovation: Building a Conceptual Framework of the Smart City through a Social Innovation Perspective," *Cities*, vol. 120, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[2] Wenwen Li, Michael Batty, and Michael F. Goodchild, "Real-Time GIS for Smart Cities," *International Journal of Geographical Information Science*, vol. 34, no. 2, pp. 311-324, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[3] Ahmed Hassebo, and Mohamed Tealab, "Global Models of Smart Cities and Potential IoT Applications: A Review," *IoT*, vol. 4, no. 3, pp. 366-411, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[4] Md Khurram Monir Rabby, Muhammad Mobaidul Islam, and Salman Monowar Imon, "A Review of IoT Application in a Smart Traffic Management System," *2019 5th International Conference on Advances in Electrical Engineering*, Dhaka, Bangladesh, pp. 280-285, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[5] Gregory Trencher, and Andrew Karvonen, *Stretching "Smart": Advancing Health and Well-being through the Smart City Agenda*, 1st ed., Smart and Sustainable Cities?, Routledge, pp. 54-71, 2020. [Google Scholar] [Publisher Link]

[6] Kang Gao, and Yijun Yuan, "Is the Sky of Smart City Bluer? Evidence from Satellite Monitoring Data," *Journal of Environmental Management*, vol. 317, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[7] Sebastian Molinillo et al., "Smart City Communication via Social Media: Analysing Residents' and Visitors' Engagement," *Cities*, vol. 94, pp. 247-255, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[8] Zaib Ullah et al., "Applications of Artificial Intelligence and Machine Learning in Smart Cities," *Computer Communications*, vol. 154, pp. 313-323, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[9] Jalal Safari Bazargani, Abolghasem Sadeghi-Niaraki, and Soo-Mi Choi, "A Survey of GIS and IoT Integration: Applications and Architecture," *Applied Sciences*, vol. 11, no. 21, pp. 1-23, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[10] Nawel Lafioune, and Michèle St-Jacques, "Towards the Creation of a Searchable 3D Smart City Model," *Innovation & Management Review*, vol. 17, no. 3, pp. 285-305, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Wang Tao, "Interdisciplinary Urban GIS for Smart Cities: Advancements and Opportunities," *Geo-Spatial Information Science*, vol. 16, no. 1, pp. 25-34, 2013. [CrossRef] [Google Scholar] [Publisher Link]

[12] Mohanaed Falih et al., "Exploring the Potential of Deep Learning in Smart Grid: Addressing Power Load Prediction and System Fault Diagnosis Challenges," *AIP Conference Proceedings*, Tabriz, Iran, vol. 3092, no. 1, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[13] Ammar Mohammedali, *Human Activities Recognition in Still Image*, Lap Lambert Academic Publishing, 2016. [Google Scholar] [Publisher Link]

[14] Nibras Kadhim Abed, Arfan Shahzad, and Ammar Mohammedali, "An Improve Service Quality of Mobile Banking Using Deep Learning Method for Customer Satisfaction," *AIP Conference Proceedings*, Kuala Terengganu, Malaysia, vol. 2746, no. 1, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[15] Siddhant Ray, "A Comparative Analysis and Testing of Supervised Machine Learning Algorithms," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 1-8, 2018. [Google Scholar]

[16] Batta Mahesh, "Machine Learning Algorithms-A Review," *International Journal of Science and Research*, vol. 9, no. 1, pp. 381-386, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[17] George Obaido et al., "Supervised Machine Learning in Drug Discovery and Development: Algorithms, Applications, Challenges, and Prospects," *Machine Learning with Applications*, vol. 17, pp. 1-20, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[18] Someah Alangari, "An Unsupervised Machine Learning Algorithm for Attack and Anomaly Detection in IoT Sensors," *Wireless Personal Communications*, pp. 1-25, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[19] Paul Kiyambu Mvula et al., "A Survey on the Applications of Semi-Supervised Learning to Cyber-Security," *ACM Computing Surveys*, vol. 56, no. 10, pp. 1-41, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[20] Nasim Tohidi, and Rustam B. Rustamov, *A Review of the Machine Learning in GIS for Megacities Application*, Geographic Information Systems in Geospatial Intelligence, pp. 29-53, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[21] Anthony G.O. Yeh, "From Urban Modelling, GIS, the Digital, Intelligent, and the Smart city to the Digital Twin City with AI," *Environment and Planning B: Urban Analytics and City Science*, vol. 51, no. 5, pp. 1085-1088, 2024. [CrossRef] [Google Scholar] [Publisher Link]

[22] Umesh Kumar Lilhore et al., "Design and Implementation of an ML and IoT Based Adaptive Traffic-Management System for Smart Cities," *Sensors*, vol. 22, no. 8, pp. 1-26, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[23] Abigail Francisco, Neda Mohammadi, and John E. Taylor, "Smart City Digital Twin–Enabled Energy Management: Toward Real-Time Urban Building Energy Benchmarking," *Journal of Management in Engineering*, vol. 36, no. 2, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[24] Danuta Szpilko et al., "Artificial Intelligence in the Smart City—A Literature Review," *Engineering Management in Production and Services*, vol. 15, no. 4, pp. 53-75, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[25] Patrycja Szarek-Iwaniuk, and Adam Senetra, "Access to ICT in Poland and the Co-Creation of Urban Space in the Process of Modern Social Participation in a Smart City—A Case Study," *Sustainability*, vol. 12, no. 5, pp. 1-21, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[26] Riadh AL-Dabbagh, "Dubai, the Sustainable, Smart City," *Renewable Energy and Environmental Sustainability*, vol. 7, pp. 1-12, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[27] Taruna Bansal et al., "Gender and Smart City: Canvassing (in)Security in Delhi," *GeoJournal*, vol. 87, pp. 2307-2325, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[28] Habib M. Alshuwaikhat, Yusuf A. Adenle, and Thamer Almuhaidib, "A Lifecycle-Based Smart Sustainable City Strategic Framework for Realizing Smart and Sustainability Initiatives in Riyadh City," *Sustainability*, vol. 14, no. 14, pp. 1-17, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[29] François Mancebo, "Smart City Strategies: Time to Involve People. Comparing Amsterdam, Barcelona and Paris," *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, vol. 13, no. 2, pp. 133-152, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[30] Afnan Alotaibi et al., "Kingdom of Saudi Arabia: Era of Smart Cities," *2022 2nd International Conference on Computing and Information Technology*, Tabuk, Saudi Arabia, pp. 285-292, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[31] Yongchang Zhang et al., "Big Data and Artificial Intelligence Based Early Risk Warning System of Fire Hazard for Smart Cities," *Sustainable Energy Technologies and Assessments*, vol. 45, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[32] Daniel Luckey et al., "Artificial Intelligence Techniques for Smart City Applications," *Proceedings of the 18th International Conference on Computing in Civil and Building Engineering*, pp. 3-15, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[33] Ahmad J. Showail, "Solving Hajj and Umrah Challenges Using Information and Communication Technology: A Survey," *IEEE Access*, vol. 10, pp. 75404-75427, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[34] Jong-Sung Hwang, *The Evolution of Smart City in South Korea: The Smart City Winter and the City-as-a-Platform*, Smart Cities in Asia, Edward Elgar Publishing, pp. 78-92, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[35] Kenneth Li-Minn Ang et al., "Emerging Technologies for Smart Cities' Transportation: Geo-Information, Data Analytics and Machine Learning Approaches," *ISPRS International Journal of Geo-Information*, vol. 11, no. 2, pp. 1-45, 2022. [CrossRef] [Google Scholar] [Publisher Link]