Chapter – 1
Introduction

The Internet of Things is a technology-based system that uses numerous devices and gadgets in place of human contact. This enables the emergence of smart cities everywhere in the world. The development of smart city systems for sustainable living, improved comfort, and enhanced productivity for people has been accelerated by the internet of things, which hosts several technologies and permits interactions between them. The Internet of Things for Smart Cities has a significant impact on a wide range of businesses, and it relies on a variety of underlying technologies to function. In this research, the Internet of Things in Smart Cities is fully studied. The key components of the IoT[1]-based Smart City environment are given first, followed by the architectures, networking, and Artificial Algorithms that enable these domains in IoT-based Smart City systems.

1.1 Smart City

The world is changing quickly, cities are expanding, and the urban population is growing. Housing affordability pushes the public to live far from their places of employment, which increases the demand for transportation and exacerbates commuting issues. The world's future lies in smart cities, which will exhibit human-like agility and sharp reasoning. Building smart cities will contribute more to fostering economic development. Any city's ability to grow successfully and sustainably depends on mobility and transportation. We must increase mobility, but we must do so more intelligently.

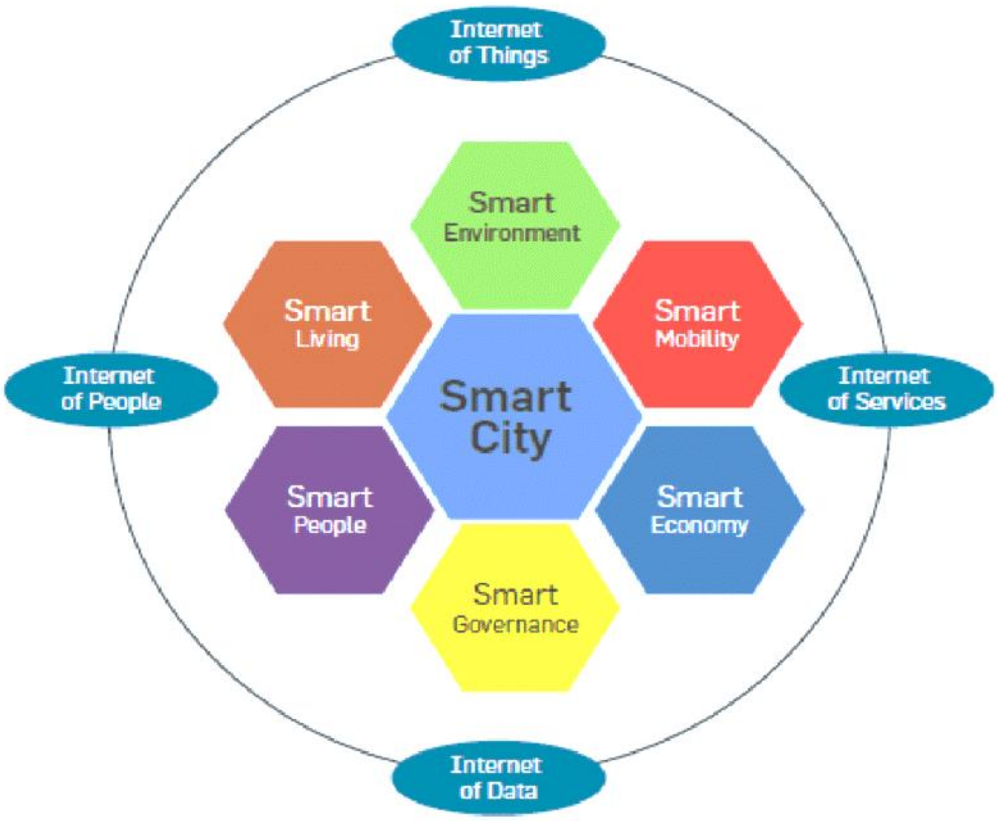
A cleverly organized network of interconnected gadgets and objects—also alluded to as a "Advanced Digital City"—that exchange information by means of wireless innovations and the cloud make up a noteworthy portion of this ICT[2] system. Real-time information collection, examination, and administration capabilities given by cloud-based IoT apps help businesses, governments, and people in taking more brilliant choices that upgrade their quality of life. Smartphones, portable gadgets, associated automobiles and homes are a couple of the ways that citizens associate with the ecosystems of intelligent cities. Costs can be decreased and sustainability can be expanded by integrating gadgets and information with a city's physical foundation and administration. With the utilization of the IoT, communities can distribute energy more

[1] Internet of Things
[2] Information Communication and Technology

effectively, collect waste more effectively, relieve traffic blockage, and upgrade Air quality. AQI[3] is a measure of how polluted the air is, or is likely to be. Problems associated with high and low AQI levels include: Health effects: Respiratory diseases: High AQI levels can cause respiratory diseases such as asthma, bronchitis, and more. may worsen conditions such as respiratory illness. Cardiovascular effects , Environmental Impact, Climate Change and Economic cost:

The lack of suitable public transportation is the fundamental problem with urban mobility. Users feel uncomfortable while the system responds to a temporary rise in demand due to increase of crowd. By utilizing smart commuting, we can increase accessibility for those who use public transportation.

Figure 1.1: Smart City Integration



The figure above shows the integrated environment required to develop smart city, Six important components are listed in the figure 1.1. The integration includes the following aspects:

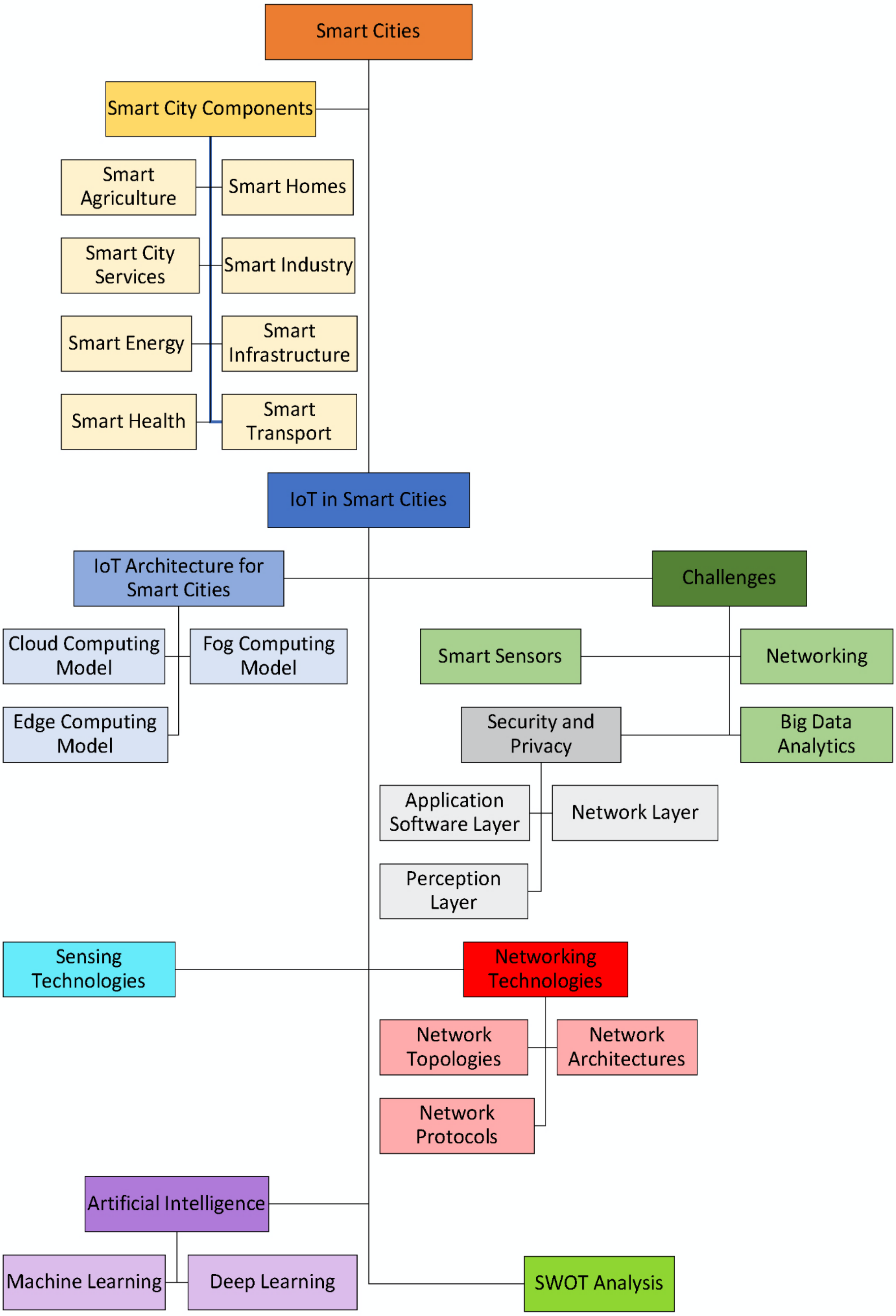
[3] Air Quality Index

- Smart Environment
- Smart Mobility
- Smart Economy
- Smart Governance
- Smart People
- Smart Living
- Internet of Things
- Internet of Services
- Internet of Data
- Internet of People

Smart mobility is an important part of the broader concept of smart cities and represents a paradigm shift in urban planning and transportation. As cities grapple with challenges such as population growth, environmental sustainability, and the need for efficient transportation, the integration of smart mobility solutions is becoming increasingly important. The Bus information system is the important extension of smart mobility, which can feature a bus location information subsystem that gathers bus data and transfers it to a traffic information centre where users can get real-time bus-related information on their mobile phones. Information such as bus operating routes, bus locations, seat availability, arrival times, and bus interval periods boosts public transportation in smart cities and increases consumer happiness.

There is abundant literature available and research is done in smart cities which cover a wide topics related to smart city components and provide insight into the technical, social, and governance aspects that contribute to smart city development and sustainability. Comprehensive understanding of the various factors are required for shaping the future of urban environments. The smart city concept incorporates different components that contribute to the, by and large, goal of making a more productive, economical, and habitable city environment. The following diagram shows the various components of smart cities.

Figure 1.2: Smart City Components & IoT



Source: Syed and Kumar (2021)

Research into the integration of smart city components and the IoT has unlocked a wide range of possibilities and redefined the way we conceptualize, plan, and live in cities. Looking back at the interconnected web of technology, infrastructure, and data that underpins smart cities, it is clear that our urban environments are on the brink of transformation. The combination of smart components and his IoT has the potential to address pressing urban challenges and promote sustainability, efficiency, and improved quality of life for citizens.

1.2 Transportation Problems in Smart Cities

The main challenges associated with smart city transportation are traffic congestion, bus location, seat availability, arrival times, and bus interval time.

1.2.1 Traffic Congestion

Traffic congestion is an important aspect of building smart and sustainable cities. If the problem of traffic congestion is not properly addressed, it can affect various aspects of urban life, economy, environment, and public welfare, resulting in several negative effects. The possible consequences of not addressing traffic congestion are:

Increased Travel Time: Congestion slows traffic flow and increases travel time. Commuters experience delays, lost productivity, and reduced overall transportation efficiency.

Economic Impact: Economic costs associated with traffic congestion, reduce productivity and increased fuel consumption and increases the ecological footprint of transportation, contributing to climate change and environmental degradation.

Environmental Impact: Air pollution and greenhouse gas emissions are the main causes of traffic congestion in cities leading to serious health effects due to long-term exposure to traffic-related pollutants. Stop-and-go traffic contributes to increased emissions and poor air quality. Businesses experience increased transportation costs and the economy as a whole becomes less efficient.

Public Transportation Efficiency: Congestion affects the reliability and efficiency of public transportation. Bus delays may occur, making public transport less attractive and less competitive with private cars.

Impact on Emergency Services: Congestion can affect emergency services response times. Delays in reaching an emergency can have serious consequences and impact public safety.

Stress and Health Problems: Commuters suffer from stress due to delays caused by traffic jams. Increased stress causes psychological problems and reduces the overall quality of life of city dwellers.

Accessibility Limitations: Traffic congestion impedes accessibility and makes it difficult for people to reach their destinations on time. Accessibility limitations impact businesses, schools, and healthcare facilities, reducing the quality of life for the entire city.

Urban Planning Challenge: Traffic congestion makes urban planning and infrastructure development difficult. Cities face the challenge of optimizing land use, creating sustainable transportation solutions, and managing population growth.

Increase in Traffic Accidents: Traffic congestion increases the risk of traffic accidents. Increased accident rates result in injuries and fatalities, and further strain on emergency services.

Quality of Life: Traffic congestion affects the general well-being and daily life of citizens. some of the ways in which traffic congestion can affect quality of life are time wastage, health impacts, Economic cost, Environmental effect, social life etc.

Inefficiency in Moving Goods: Traffic congestion hinders the movement of goods and delivery services. Logistics and supply chains are facing delays, impacting businesses and consumer services.

Understanding these impacts highlights the importance of implementing smart and sustainable transport solutions to address urban traffic congestion. Strategies may include investing in public transport, implementing smart traffic management systems, promoting different modes of transport, and promoting urban planning that prioritizes accessibility and sustainability. Reducing traffic congestion is essential to creating liveable, sustainable and economically dynamic urban spaces. Deploying intelligent transportation solutions, investing in public transport, promoting different modes of transportation, and adopting efficient city planning strategies can help reduce the negative effects of traffic congestion.

1.2.2 Bus Location

The lack of accurate bus location information in smart cities can lead to a number of issues that impact both the efficiency of public transportation and the overall commuting experience. Here are some issues related to the lack of bus location services in smart cities:

Unsafe Commuting Experience: Without real-time bus location services, commuters are uncertain about their current location and bus arrival time. This uncertainty can lead to longer wait times at bus stops, making it difficult for commuters to plan their trips effectively.

Inefficient Route Planning: Lack of real-time location data prevents the ability to dynamically adjust bus routes based on current traffic conditions. Buses follow a static schedule, which can cause inefficiencies when traffic jams, road closures, or unforeseen events occur.

Increased Congestion: Without real-time tracking, buses can be delayed, leading to increased congestion and congestion. When routes become congested, travel times become slower, less reliable, and public transportation options become less attractive.

Low Operational Efficiency: Without accurate location data, transit agencies cannot effectively monitor and manage their bus fleets. This can lead to operational inefficiencies, difficulty resolving maintenance issues in a timely manner, and difficulty optimizing service levels.

User Accessibility Limitations: Commuters do not have access to real-time information about bus locations, making it difficult to plan and coordinate their trips. This restriction may have a disproportionate impact on people using public transport and may impact the accessibility of the transport system as a whole.

Reduced Passenger Satisfaction: Without real-time information, passengers become frustrated with unpredictable bus arrivals and potential service interruptions. Lower passenger satisfaction leads to lower public trust and may lead to less choice of public transport as a preferred mode of transport.

Missed Connections: Without real-time tracking, commuters can miss connections to other modes of transportation. Connection failures can increase travel time, stress, and compromise a seamless intermodal transportation experience.

Inefficient Traffic Management: Without bus location data, cities cannot effectively manage traffic and optimize bus signal time. This results in suboptimal traffic flow, congestion, and overall inefficiency of the transportation network.

Emergency Response Difficulties: During bus emergencies and accidents, the lack of real-time location data complicates emergency response efforts. Delays in emergency response may jeopardize public safety and exacerbate the severity of the incident.

Underutilization of Smart City Technology: Lack of bus tracking services means missed opportunities to make the most of smart city technology to improve transportation. Smart city initiatives may not reach their full potential to optimize urban mobility, reduce environmental impact, and improve overall quality of life.

Ensuring accurate bus tracking services in smart cities is critical to building an efficient, reliable, and easy-to-use public transportation system. By addressing these issues and implementing real-time tracking solutions, smart cities can improve the overall commuting experience, promote sustainable transportation, and contribute to the success of broader smart city initiatives.

1.2.3 Seat Availability

Lack of information about seat availability in smart cities can lead to various challenges and inconveniences for public transport users. Here are some issues related to the lack of availability services in smart cities.

Unsafe Commuting Experience: Commuters do not know about seat availability on public transportation. Passengers may lack confidence in finding a seat, which may cause discomfort and inconvenience during the journey.

Crowded Vehicles: Lack of availability information leads to congestion on buses and Passengers may be required to stand only, reducing overall comfort and potentially leading to safety concerns.

Inefficient Use of Space: Without real-time data on seat occupancy, a transportation company cannot optimize its use of space. Buses can have uneven passenger distribution, which can result in wasted space and operational inefficiencies.

Decreased Passenger Satisfaction: Passengers can become frustrated if they cannot find a seat. Lower satisfaction with public transport services can lead to lower ridership and negative perceptions of the overall public transport experience.

Inconvenience for Passengers with Special Needs: Passengers with special needs such as Elderly people or pregnant ladies and people with disabilities may have difficulty finding an available seat. Lack of accessibility can cause inconvenience and difficulty to passengers who rely on the seats.

Suboptimal Resource Allocation: A transportation company cannot allocate resources efficiently without real-time data about seat occupancy. Suboptimal resource allocation can lead to unnecessary operating costs and problems meeting passenger demand.

Uncomfortable Commuting Environment: Crowded vehicles can create an unpleasant commuting environment. Passengers may suffer from stress, decreased sense of well-being, and an overall negative perception of public transport.

Inefficient Public Transportation Planning: Public transportation planners do not have accurate data about seat demand. This can result in suboptimal route planning, lack of capacity to meet peak demand, and operational issues.

Customer Dissatisfaction: Passengers who cannot find an available seat may express dissatisfaction with public transportation. Negative feedback and decreased customer satisfaction can affect the reputation and attractiveness of public transport.

Challenges for Long Distance Travellers: Long-distance travellers may have difficulty finding a seat. Feeling unwell while traveling long distances may prevent people from choosing public transportation for long distance journeys.

Providing real-time availability information is important to improve the overall quality of public transport services in smart cities. Addressing these issues by implementing seat availability services will result in a more comfortable, efficient, and user-friendly transportation experience and encourage the adoption of sustainable transportation options.

1.2.4 Bus Interval Time

Bus interval time, or the time between consecutive bus arrivals, is an important aspect of public transportation in smart cities. Issues with bus spacing can affect the efficiency and attractiveness of bus services. Some issues related to bus interval times in smart are

Irregular Bus Intervals: Inconsistency in bus arrival times. Passengers may experience unexpected waiting times, causing inconvenience and frustration.

Overcrowded Buses: Insufficient spacing can lead to overcrowded buses. Passenger discomfort, safety concerns, and negative impact on the overall quality of the transportation experience.

Reduced Reliability: Unpredictable bus intervals reduce the reliability of public transportation. Commuters may be reluctant to rely on buses for their daily commute and may choose alternative modes of transportation.

Making Public Transportation Less Attractive: Inefficient bus spacing makes public transportation less attractive. Cities may face challenges in promoting sustainable transport, leading to increased reliance on private cars.

Increased Travel Time: The longer the bus interval, the longer the travel time. Long travel times can deter people from using public transportation, especially for time-sensitive activities.

Operational Inefficiency: Poor management of bus spacing can result in operational inefficiency. Transportation operators may have difficulty optimizing their resources, increase operating costs, and have difficulty meeting passenger demand.

Difficulty in Timed Transfers: Irregular bus intervals make it difficult for passengers to make timed transfers between different bus routes. Commuters may miss their connections, causing overall travel disruption.

Unequal Service Distribution: Bus spacing can be unevenly distributed across different routes or districts. Some regions may experience longer distances, creating disparities in service quality and accessibility.

Implications for Urban Planning: Irregular bus spacing can complicate urban planning efforts. Cities may face challenges in optimizing land use, designing efficient transportation corridors, and creating sustainable urban environments.

Negative Environmental Impact: Inefficient bus spacing increases consumption of fuel consumption and produces carbon emissions. The environmental footprint of public transport is increasing, impacting air quality and sustainability goals.

Frequency Modal Shift: People cannot switch to public transportation because bus intervals are unreliable. Cities may struggle to reduce traffic congestion and promote more sustainable transportation options. Irregular bus intervals can force people to take alternate transport arrangement which results in revenue loss.

Activity Planning Difficulty: Unpredictable bus intervals make it difficult for passengers to plan their activities. Commuters may have difficulty scheduling appointments, work, and other daily activities.

To create an efficient, reliable, and attractive public transport system in smart cities, it is important to solve problems related to bus spacing. The implementation of smart technology, data analytics and effective scheduling strategies are key elements of the solution to optimize bus spacing and improve the overall quality of public transport services.

1.2.5 Public Addressing System

The lack of a PAS[4] in smart cities presents a variety of challenges and drawbacks that can impact the efficiency, safety, and overall user experience of transportation and public communication systems. Some of the issues related to the lack of provision of public information systems in smart cities are:

Lack of Real-Time Information: Without PAS, there will be lack of communication of real-time information to the public about traffic schedules, delays, emergencies, and other important updates. This can cause confusion for commuters and lead to inefficiencies in the transportation system.

[4] Public Addressing System

Reduced Emergency Preparedness: Public address systems are essential for emergency notification and evacuation procedures. Without these systems, it would be difficult to convey emergency information quickly and effectively, putting public safety at risk during crises such as natural disasters or security incidents.

Accessibility Limitations: PAS is essential for providing information to passengers with visual or hearing impairments and ensuring access to public transport. The lack of such systems can make it difficult for people with disabilities to use public transport and get around.

Inefficient Traffic Management: Without the ability to communicate real-time traffic updates, route changes, and alternative options to commuters, the overall efficiency of traffic management in smart cities can decrease. This may increase congestion and delays.

Poor Passenger Experience: Lack of PAS impacts the overall passenger experience as commuters are unable to obtain important information about their journey. Passengers may experience frustration, inconvenience and dissatisfaction with public transport.

Limited Public Participation: The public information system helps engage the public by providing information about City events, cultural activities, and community announcements. The lack of these systems can lead to a lack of community awareness and participation.

The Difficulty of Coordinating Multimodal Transportation: Smart cities often integrate different transportation modes, and PAS can help coordinate these transportation modes seamlessly. Without such a system, coordination between buses, trains, trams, and other modes of transportation becomes even more difficult.

Reduced Operational Efficiency: PAS contributes to the efficient operation of transportation services by providing real-time updates to drivers, maintenance personnel, and other stakeholders. Without these systems, operational efficiency may decrease and operating costs may increase.

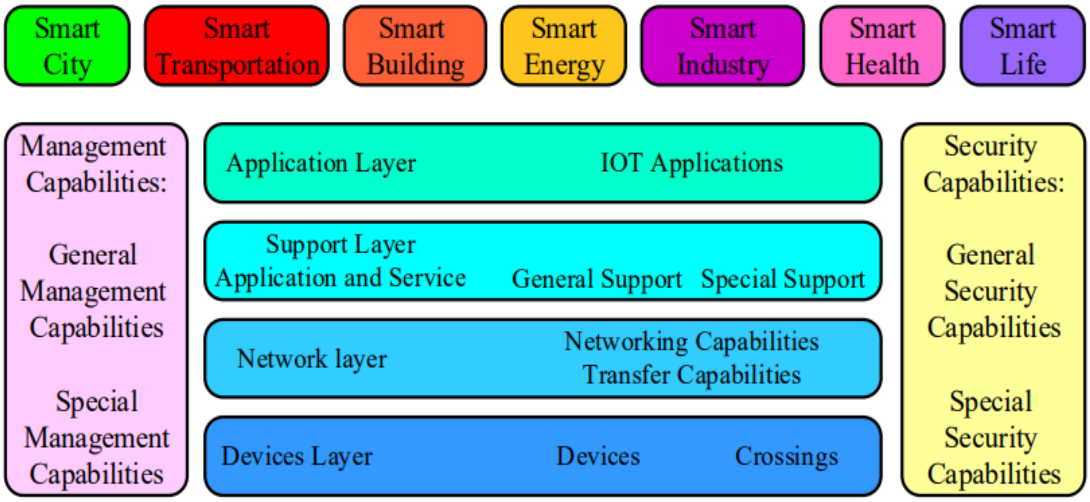
Limited Adaptability to Smart Infrastructure: In smart city environments, integration of PAS with other smart systems such as smart traffic lights and sensors is essential for optimal traffic management.

In summary, the lack of a public address system in smart cities can lead to a number of problems, including limited real-time communication, security breaches, poor accessibility, and overall inefficiency of transportation and public services.

1.3 Internet of Things

The worldwide worthiness of IOT administrations and huge information analytics have reinforced smart city ventures. These administrations have essentially moved forward the quality of human life by progressing the infrastructure and transportation framework, lessening traffic congestion, and giving waste disposal. This article presents a comprehensive audit of the integrated ICT network types, attainable potential, and critical prerequisites for the IOT worldview for smart cities. The Internet of Things is a network of physical objects, such cars, structures, and even the vital electrical appliances we use every day that are connected to one another over the internet so they may gather and share data among themselves. These "Things" can prioritize tasks, self-organize, and interact with other things without the help of humans. Each individual has more than six devices linked to the Internet. The idea behind the Internet of Things is to make the web even more pervasive and immersive. Additionally, by facilitating simple access to and interaction with a wide range of devices, for example, household appliances, monitoring, security cameras, sensors, displays, actuators, and automobiles. The Architecture of IOT is shown.

Figure 1.3: Architecture of IoT



Source: Moazzami and Majid(2021)

The IoT is characterized by a layered architecture that organizes and structures the flow of information and data. This architecture enables efficient communication and coordination between the various components of an IoT system. The hierarchical structure of IoT information provides a systematic approach to designing, implementing, and managing IoT systems. A typical hierarchical structure for IoT information is shown above, It has four layers:

Application Layer: The application layer in the context of the IoT refers to the top layer of the IoT architecture that is responsible for communication and interaction between IoT devices and applications. This layer plays a critical role in facilitating data exchange, data processing, and application-specific functionality. Important aspects of the application layer in IoT are: Data processing and analysis: The application layer processes and analyses data generated by IoT devices. The goal is to gain valuable insights from raw data in order to make informed decisions. Application-specific protocols: IoT applications often require specific protocols for communication. The application layer defines the rules and conventions for exchanging data between devices and applications to ensure interoperability. APIs[5] and Web Services: APIs and web services are used at the application layer to enable communication between the various components of the IoT ecosystem. This enables third-party developers to create applications that interact with IoT devices and data. In summary, the application layer of IoT is responsible for managing the communication, processing, and functionality that defines how IoT devices interact with applications and services.

Support Layer: In the context of IoT architecture, the support layer includes various components and functions that provide support services essential to the smooth operation of the IoT ecosystem. This layer typically includes the overall infrastructure, management, and elements that contribute to optimizing IoT deployments.

Network Layer: The network layer of the IoT architecture plays a critical role in facilitating communication and data exchange between a myriad of interconnected devices. These are the infrastructure and protocols that allow devices to connect, share information, and operate consistently within the IoT ecosystem. The key components and aspects of the network layer in an IoT architecture are: Connection Protocols: The network layer involves the selection and implementation of connection protocols that

[5] Application Program Interfaces

allow devices to communicate with each other and the broader network. Common IoT protocols include MQTT[6], CoAP[7], and HTTP[8]/HTTPS. Wireless Communication Technology: IoT devices often rely on wireless communication technology for flexibility and ease of deployment. These may include Wi-Fi, Bluetooth, Zigbee, Z-Wave, cellular networks (3G, 4G, 5G), and LPWAN[9].

Device Layer: The lowest layer of an IoT architecture is often referred to as the device layer or perception layer. This layer is the foundation of the IoT ecosystem and consists of physical devices or "things" that collect data, sense the environment, and interact with the real world. Devices in this layer are responsible for collecting information and passing it to higher layers for processing and analysis. The key components and aspects of the lowest layer of the IoT architecture are: Sensors and Actuators: Sensors are devices that detect and measure physical properties such as temperature, humidity, pressure, light, and movement. Actuators, on the other hand, allow devices to perform actions based on the data they receive, such as turning on/off, adjusting settings, or triggering other physical processes. IoT devices are often embedded systems with embedded computing capabilities. These systems can process data locally, perform simple calculations, and control connected sensors and actuators. Microcontrollers and microprocessors provide the computing power needed to operate IoT devices. They control device functionality, handle data processing, and manage communication with other devices or networks. The lowest level device is equipped with a communication module that allows it to send and receive data. The data collected by devices at this layer is transferred to the network layer for further processing, analysis, and decision-making.

Internet of Things architecture is critical to the effective deployment and functioning of IoT systems. It provides a structured framework for connecting and managing a variety of devices and sensors, enabling seamless communication, data sharing, and intelligent decision-making.

[6] Message Queuing Telemetry Transport
[7] Constrained Application Protocol
[8] Hypertext transfer protocol
[9] low-power wide area network

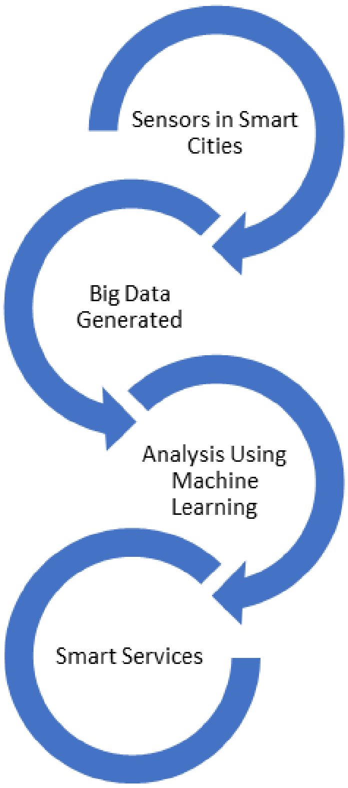
1.4 Role of Machine Learning

It is common knowledge that many individuals would relocate from rural to urban areas as a result of industrialization. The transition from rural to urban areas will bring about a number of challenging issues, including increased traffic, pollution, stress on the fundamental infrastructure, and challenges with waste management. Finding effective, ethical, and sustainable solutions to improve lives will become increasingly important as a result. This is where the still-evolving idea of a "smart city" comes into play. SCs[10] are designed to effectively manage energy use, preserve the environment, raise the economic and living standards of their residents, and improve their capacity to effectively employ and adopt contemporary information and communication technologies. SCs have cutting-edge sensors that manage city resources and contribute significantly to the gathering of crucial data that can subsequently be used in a variety of applications. The different types of sensors include electronic ones like parking and speedometer sensors, chemical ones like oxygen and carbon dioxide sensors and catalytic bead sensors, biosensors for identifying biomedical components, and smart grid sensors for effective power generation, transmission, and distribution from the point of generation to the users. Most of the data flowing from the sensors is wasted since there are no set standards or processes for extracting potentially useful information. To analyse massive data as effectively as feasible, consider long-term goals, and arrive at the best or nearly best conclusions, machine learning (ML[11]) can be utilized.

A subfield of AI[12], ML enables computers to automatically learn by exploring the data made available and using it to gain experience and learn new things without being explicitly programmed. Arthur Samuel first used the phrase "Machine Learning" in 1959. Since the development of the first artificial neural network for computers, the perceptron, and the use of the "nearest neighbor" algorithm to solve the nearest neighbour problem, ML has advanced from being a simple tool for pattern recognition to one that can carry out complex tasks and incorporate deeper domains of AI.

[10] Smart Cities
[11] Machine Learning
[12] Artificial Intelligence

Figure 1.4: Steps of Applying Machine Learning Techniques



A system that can learn from past experiences and develop on its own is required due to the abundance of data generated by sensors and smart devices from SCs that cannot be inspected individually. As a smart city application's context and operational environment change, it need a generic, dynamic, and ongoing learning mechanism. Therefore, it is essential for increased efficiency to investigate the possibilities of ML and big data in the creation of individualized services in SCs. The ML technique is rapidly changing due to progressions in calculations, information collection procedures, computer networks, modern sensors, and IO gadgets, and interest in self-customization to client conduct.

The primary goal of machine learning algorithms is to accurately understand data that has never been "seen" before and make predictions that go beyond the training samples, similar to data from the real world. By increasing the amount of training data and strengthening their ability to learn, the algorithms' accuracy and precision can be further improved. Based on the type of learning "signal" or "feedback" that a learning system is able to receive, machine learning algorithms can be roughly divided into three groups.

1.4.1 Supervised

Giving training data that has previously been "known" or "labelled" with the proper response and consists of N input-output pairs (X, Y) is how supervised learning functions. The ANN[13] then generates an output Z for each unknown X, which is then compared against Y using an error (cost or distance) function. Finally, an iterative process is used to minimize this mistake. Image Classification: Training with image/label datasets are examples of supervised learning methods. A new image is then presented later with the hope that the computer will pick up on the new object. Regression: Giving the system marked historical data so it can forecast the future result of an identical circumstance.

1.4.2 Unsupervised

Using unsupervised learning methods, it self-organizes and finds hidden patterns in unlabelled input data to create neural networks. It can analyse data without sending an error signal so that the potential fix can be assessed. Unsupervised learning can occasionally be useful since it allows the algorithm to search the past for patterns that weren't previously taken into account. Unsupervised learning is necessary because manually inspecting huge datasets like those for speech recognition is highly expensive. Clustering is a very basic but well-known example of unsupervised learning.

1.4.3 Semi Supervised

This category is a hybrid of the previous two. The algorithm is trained on a dataset that contains both labelled and unlabelled data. It works by taking enormous amounts of input data and labelling only a subset of it as training data. Reinforcement learning, a related strategy, provides feedback to guide the computer programme in interacting with a dynamic environment. In this approach, a model is deployed using a small set of labelled samples and a larger set of unlabelled samples. The goal is to use labelled data to make predictions about unlabelled data and use the additional information to improve model performance.

[13] Artificial Neural Network

1.5 Machine Learning Algorithms

To avoid imprecise or erroneous predictions, the data collected/generated must go through pre-processing, merging, modifying, and cleaning (removing null values). The computational intensity and speed of a specific technique are two significant characteristics to consider while employing ML techniques. The best algorithm is chosen based on the user application and should be fast enough to track changes in the input data and provide the desired output in a reasonable amount of time. ML algorithms create a mathematical model using sample data, known as "training data," on which to make predictions or choices. The training phase of supervised ML classifier development involves training a specific classifier from a set of labelled data. As the size of the training data increases, so does the performance of the classifiers. Some of the most popular ML algorithms are detailed further below.

1.5.1 Bayes Net

A Bayesian network, also known as a Bayes net or belief network, is a graphical model that represents probabilistic relationships between a set of variables. Bayes net are not algorithms themselves, but they serve as a framework for thinking under uncertainty. Several algorithms are concerned with learning the structure and parameters of Bayes net from data. The two main types of algorithms related to Bayes net are:

Structure Learning Algorithms: These algorithms determine the graphical structure of Bayes net and specify dependencies and conditional relationships between variables.

Parameter Learning Algorithm: Once the structure of the Bayes net is known, the parameter learning algorithm estimates the parameters (conditional probability distributions) associated with each node based on its parents in the network.

1.5.2 Naive Bayes

Based on Bayes' theorem, a naive Bayes classifier is a probabilistic classifier that works by assuming that no pair of features are dependent. Naive Bayes is a simple but powerful machine learning algorithm based on Bayes' theorem and the assumption of independence between features. Despite its simplicity, Naive Bayes is often effective and computationally efficient, so it is often used in a variety of classification tasks. It is particularly suitable for text classification and spam filtering.

1.5.3 Logistic

Logistic regression is a machine learning algorithm commonly used for binary classification tasks, where the goal is to predict whether an instance belongs to one of two classes. Despite its name, logistic regression is more of a classification algorithm than a regression algorithm. Logistic regression is a fundamental machine learning algorithm that is widely used in various applications such as medical diagnostics, spam detection, and credit scoring due to its simplicity, interpretability, and effectiveness. Although it is designed for binary classification, it can be extended to handle multiple classes through techniques such as one-vs-rest regression and softmax regression.

1.5.4 SMO

SMO[14] is a machine learning algorithm designed to train SVMs[15] in supervised learning. SVM is used for classification and regression tasks, and SMO is a specific algorithm used to efficiently solve the optimization problems associated with training these models. Although SMO is an important algorithm for SVM training, there are alternative approaches and optimizations to solve SVM problems, such as the widely used libsvm library that implements more general optimization techniques. Still, understanding SMO provides insight into the support vector machine training process.

1.5.5 IBK

IBk[16] is a machine learning algorithm used for classification and regression tasks. It is the part of the family of k-NN[17] algorithms, where the prediction of a new instance is based on the majority class for classification or mean for regression of the k-nearest neighbors in a function space. The main features of the IBk algorithm is Instance-based learning. This means that no explicit model is created during training. Instead, save the training instance and use it to make predictions for new instances. In K-NN Predictions for new instances are determined by examining the class labels for classification or values for regression of the k-nearest neighbors in the training data set. Small values of k gives the model that is more flexible and sensitive to noise, and large values of k gives the model that is smoother and less sensitive. Regression uses the average of the

[14] Sequential Minimal Optimization
[15] Support Vector Machines
[16] Instance-Based k-Nearest Neighbors
[17] k-Nearest Neighbors

k nearest neighbor target values as the prediction. IBk can be computationally expensive, especially for large datasets, as it must calculate the distance for each prediction. It is often more efficient when the dataset is small. IBk performance can be sensitive to feature scaling. Therefore, it is often recommended to normalize or standardize features to obtain a similar scale. IBk is a simple but effective algorithm, especially in situations where the decision boundary is complex and not easily captured by parametric models. It is widely used in various fields such as pattern recognition, classification, and regression.

1.5.6 K Star

K Star was developed in 2009. K Star was originally implemented as part of DiPro toolset for generating counter examples in probabilistic model checking. K Star A directed search algorithm also called as K*. It Finds the k shortest paths between the given pair of vertices in the given directed weighted graph. K Star works on the fly. This means that the graph does not have to be made explicitly available and stored in main memory. K Star can be also be controlled using a heuristic function.

1.5.7 Multi Class Classifier

A multiclass classifier is a type of machine learning algorithm that can assign instances to one of three or more classes. Unlike binary classifiers, which distinguish between two classes (such as positive or negative), multiclass classifiers handle scenarios where there are multiple possible classes. Some of the common Multi Class algorithms are Support vector machine, Random Forest, K Nearest Neighbours, Neural Networks and Decision Trees. The choice of algorithm often depends on factors such as the size and type of the dataset, computational efficiency, and the desired interpretability of the model.

1.5.8 Random Forest

A decision tree-based supervised machine learning approach called RF[18] depends on values from a random vector that is sampled separately and with the same distribution across all of the trees in a forest. By averaging the results, this ensemble method lowers over-fitting and bias-related error, leading to superior outcomes. Random Forest is a powerful and versatile machine learning algorithm that belongs to the ensemble learning category. Ensemble learning combines the predictions of multiple models to create a more robust and accurate model. Random forests are particularly effective for both classification and regression tasks. The main features and characteristics of the Random Forest algorithm are: Ensemble of Decision Trees: Random Forest is an ensemble of Decision Trees. A decision tree is a discrete model that makes predictions based on a series of hierarchical decisions. Random forests create multiple decision trees and combine their predictions during the training phase. During the training process, Random Forest randomly selects a subset of the training data (with permutations) to train each decision tree. This process is called bootstrapping. Additionally, at each decision point in the tree, a random subset of features is also considered. Random Forest uses a technique called bagging, where each decision tree is trained independently on a different subset of the data. The final prediction is determined by aggregating the predictions of all trees . By training multiple decision trees on different subsets of data and features, random forests become more robust and less prone to overfitting compared to a single decision tree. Overfitting occurs when a model learns the training data well enough but is unable to generalize to new, unseen data. Random Forest provides a measure of feature importance. Analysing the contribution of each feature across multiple trees can help determine which features have the greatest impact on predictions. The training of individual decision trees in a random forest can be performed in parallel, resulting in a scalable algorithm that can efficiently process large amounts of data. Random forests tend to be less sensitive to outliers in a dataset. Because each tree is trained on a subset of the data, the impact of outliers is reduced. Random Forest has been implemented in various machine learning libraries such as Scikit-Learn in Python, making it highly accessible and widely used.

[18] Random Forest

1.5.9 Random Tree

Random Tree is a term often associated with two different machine learning algorithms, Random Forest and Highly Randomized Trees (Extra Trees). Both algorithms fall into the category of ensemble learning and are used for classification and regression tasks. Both Random Forest and Extra Trees are powerful algorithms that leverage the concept of ensemble learning to improve predictive performance. They are widely used in various applications such as classification, regression, and feature importance analysis. The choice between random forests and extra trees may depend on the specific properties of your data and the desired trade-offs between computational efficiency and model accuracy.

1.6 Applications of Machine Learning

Machine learning algorithms are applied in a variety of fields to provide solutions to complex problems and make data-driven predictions. Here are some notable applications of machine learning algorithms:

1.6.1 Smart Traffic Management and Transportation

Traffic bottlenecks disrupt people's lives on a daily basis. It results in lengthier trips due to traffic congestion, more pollution, and huge economic losses due to delays and other transportation issues. Thus, one of the primary problems in modern cities is Smart Mobility, which focuses on providing sustainable transportation systems and logistics to allow for seamless urban traffic and commuting through the use of mostly information and communication technology. They also include ways that employ personal information to make useful recommendations for small-scale personal management tasks such as looking for free parking. To decide control actions, certain traditional control approaches such as SFC[19] and traditional AI techniques based on historical data (recorded similar occurrences) such as case-based reasoning and rule-based systems were developed.

[19] Static Feedback Control

However, these technologies have shortcomings such as difficulty dealing with traffic network dynamics and the lack of a learning mechanism to cope with unknown events and automatically update their model. The advancement of ML and DL[20] cleared the way for a generic and adaptable method of developing intelligent and adaptive traffic control systems.

1.6.2 Combating Pollution

Today, worry over pollution is on the rise. It is a significant threat to the environment as well as a primary risk factor for a number of illnesses like lung and skin diseases. Large, intricate, and varied time-series data have been subjected to a number of Machine Learning approaches during the past ten years to produce estimates for air pollution. The conventional method for predicting air quality utilized mathematical and statistical methodologies, where data was coded using differential equations and a physical model. These methods were time-consuming and only partially accurate because they couldn't foresee the extremes, such as the pollution maximum and minimum cut-offs.

1.6.3 Public Safety

Massive amounts of data are being collected in SCs by sophisticated sensor networks, which has increased the difficulty of protecting that data from intrusions and made the implementation of effective cybersecurity measures necessary. SCs are extremely susceptible to online risks including data theft and illegal device access. By employing techniques like facial and speech recognition, recognizing criminal patterns around the city to create better public policies, and detecting fraud, ML can also be utilized to facilitate the job of law enforcement authorities.

1.6.4 Smart Grids and Machine Learning

Energy management and conservation are regular 24-hour concerns for consumers and energy service providers. With the aid of digital communications technology, a smart grid is an energy network that enables the two-way flow of data and electricity, allowing us to detect and respond to changes in demand as well as a number of other difficulties. It guarantees effective electricity transmission, prompt power restoration during power

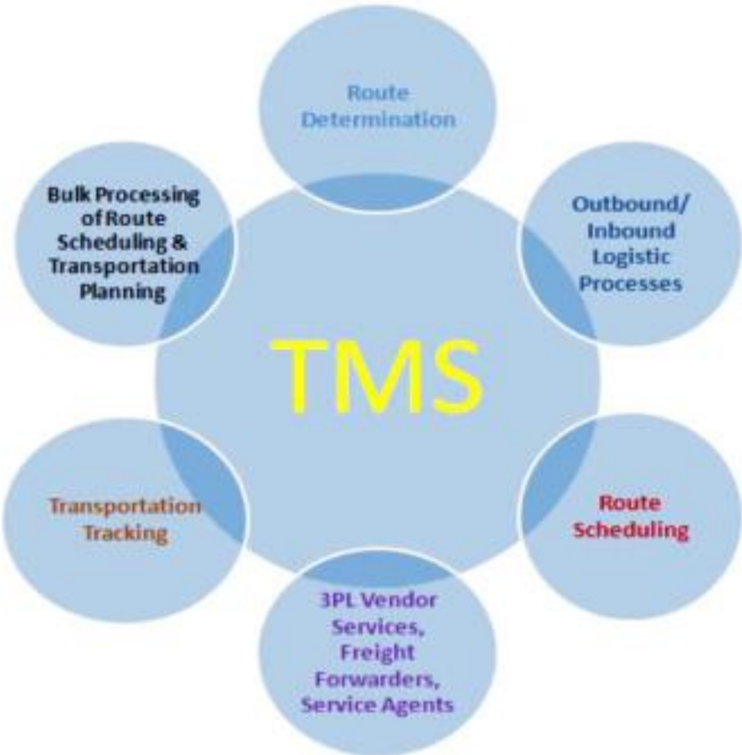
[20] Deep Learning

outages, increased customer-owner power generation system integration, including the integration of renewable energy systems, and enhanced security against electricity theft. Applying ML techniques also strengthens the power system's endurance and adaptability and improves its readiness to handle emergencies like natural and man-made disasters.

1.7 Role of Artificial Intelligence

The field of transportation has benefited greatly and substantially from AI. The solutions provide vehicle and driver safety through the use of autonomous vehicles, traffic control, optimal routing, and logistics. The data produced by the devices deployed in vehicles employing AI technology is used to build ITS[21]. Four transportation-related subsystems—the Intelligent Traffic Management System, the Intelligent Public Transport System, the Intelligent Safety Management System, and the Intelligent Manufacturing & Logistics System—are the focus of the current study.

Figure 1.5: Old Transportation System,



Source: Iyer (2021)

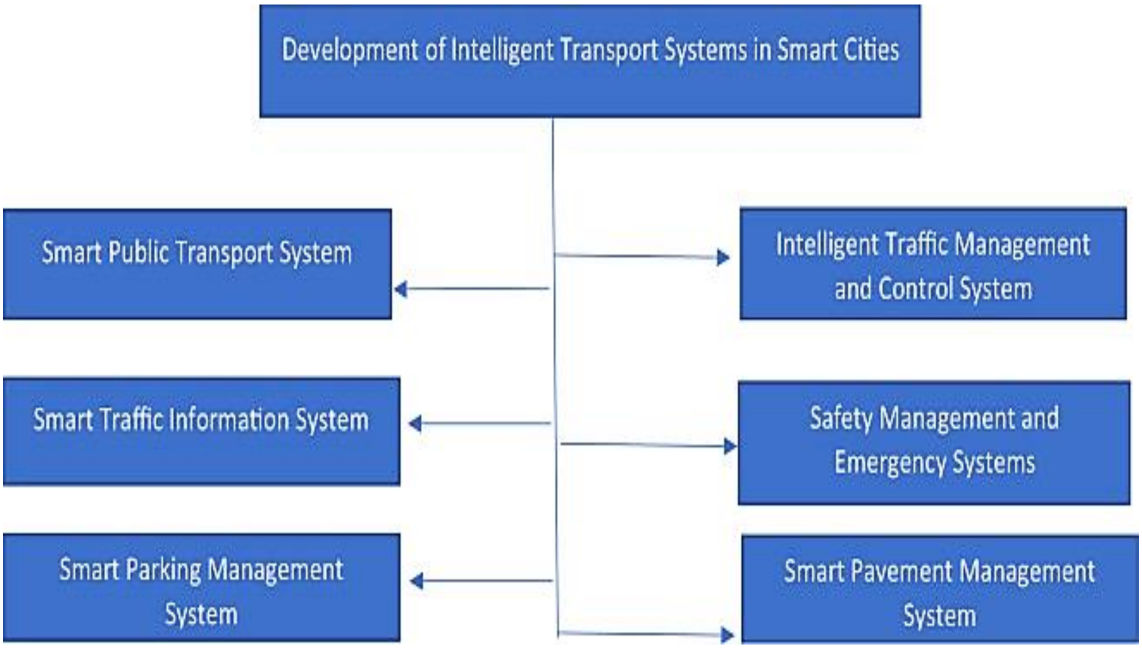
[21] Intelligent Transportation System

Old traditional transportation systems refer to the transportation methods and infrastructure that were prevalent before the advent of modern, technologically advanced transportation. These systems varied by region and historical period and were characterized by their dependence on traditional techniques and means of propulsion.

The Above figure shows old TMS[22] Where Every thing was controlled manually right from fleet management, Route determination, Route Scheduling and Inbound outbound logistics. Manual methods are time consuming, in efficient and prone to errors. Before advanced technology became widespread, such systems relied on manual processes and simple tools, such as Manual record keeping, Schedule maintenance, Fuel management, Driver management, Fixed routes and Time table, Paper schedules, Manual Adjustments.

The transition from traditional fleet management and route planning to modern ones has significantly improved efficiency, responsiveness, and overall transportation services.

Figure 1.6: Sub-systems in Intelligent Transportation System



Source: Agarwal, Gurjar and Birla (2015)

[22] Transport Management System

The above figure shows the development steps of Intelligent Transport Systems in Smart Cities. There are six important components:

Smart Public Transport System: Smart public transportation systems use advanced technology and data-driven solutions to improve the efficiency, reliability, and user experience of public transportation services. The aim is to make urban mobility more sustainable, accessible and comfortable for passengers, while optimizing the operation of public transport networks. The main components and features of an smart public transportation system are:

- Real-time tracking and monitoring
- GPS location tracking of public transport (buses, trains).
- Predictive Analytics and Arrival Time Estimation
- Contactless Ticketing and Payments
- Automatic Fare Collection

Deploying smart public transportation systems requires collaboration between transportation authorities, technology providers, and communities. Integrating new technologies, data analytics, and user-centred design principles will play a key role in transforming traditional public transport into more efficient and passenger-friendly systems.

Intelligent Traffic Management and Control System: It has been famous that machine learning algorithms are regularly utilized to figure out traffic jams and arrange routes. A city-by-city survey of the selection of AI to address transportation concerns shows that the larger part of developed countries have rapidly grasped these innovations. Given that it incorporates venture and long-term considering the portion of the best administration, this selection requires the support of the specific organizations and the administration. Due to two factors, certain businesses and governments are still on the fence about using AI: either they are worried about the risks involved, or the general level of technology adoption in these nations is low.

AI helps with transportation problems by recommending alternate routes and monitoring traffic signals in real-time amid heavy traffic. Through effective traffic management, environmental pollution can be reduced, and sustainable cities can be built. AI offers arrangements for foreseeing weather and traffic designs, managing streets, and alarming on-duty police officers. Sometime recently beginning their trip, these frameworks back drivers, commuters, and people on foot. It is vital to have innovative support to form a viable open transportation framework that helps in arranging and decision-making.

The frequency of accidents on the roads has decreased thanks to AI, which also warns drivers about road safety and anticipates accidents based on the state of the roads. When the transportation sector is productive, an economy may function successfully. It is important to accomplish this by developing secure transportation systems with the use of AI technologies. Using AI during the automobile production process For better results, sensors, cameras, and other technology have been used in this sector. Some of the automaker's built-in AI technologies are already standard equipment for both commercial and passenger vehicles.

Smart Traffic Information: It uses technology and data to optimize traffic flow, improve road safety, and provide real-time information to both drivers and traffic authorities. These systems use a variety of technologies and data sources to improve overall transportation efficiency. The main components and features of an smart Traffic information are:

- Interactive displays
- Announcements on public addressing system.
- Provide passengers with real-time information, service updates, and emergency notifications.

Smart traffic information systems play a key role in improving mobility across cities, reducing congestion, and increasing the safety and efficiency of transportation networks. Integration of technology, data analytics, and communication systems is key to the success of these smart traffic management solutions.

Safety Management and Emergency Systems: Safety management and emergency systems are an integral part of intelligent transportation solutions and contribute to the security and resilience of the entire transportation network. The main aspects and features of security management and emergency systems in intelligent transportation are:

- Real-time monitoring and monitoring
- Incident Detection and Response
- Emergency Communication Systems
- Integrated Emergency Services like police, fire, medical
- Dynamic Signs and Warnings
- Intelligent Traffic Signal prioritization for Emergency Vehicle like Ambulance and VIP Vehicles.
- Surveillance Drone Technology equipped with cameras and sensor to provide real-time surveillance and assessment, especially in hard to reach areas.
- Implement robust cybersecurity measures to protect transportation systems from cyber threats that can compromise security and disrupt operations.
- Encourage the community to actively participate in safety reporting through mobile apps, social media, or other channels.
- Provide a platform for citizens to report incidents and dangers.
- Developing evacuation plans and conduct simulations to ensure effective response and evacuation procedures during emergencies and natural disasters.

Integrating safety management and emergency systems with intelligent transportation technologies makes urban transportation networks more resilient and responsive, contributing to safer and more secure mobility for everyone.

Smart Parking Management System: Smart parking management systems use technology to improve the efficiency and user experience of parking facilities. These systems use a variety of technologies and data-driven solutions to optimize parking space utilization, reduce congestion, and improve the overall parking experience. The key components and features of smart parking management are:

- Real-time Parking Availability Using sensors and cameras.
- Parking Guide Mobile App to provides real time parking information.
- Smart Parking Meters to accept electronic payments and mobile payments.
- License Plate Recognition for vehicle identification and payment.
- Reservation System to Allow users to reserve parking spaces in advance.
- Dynamic Pricing Models based on real time demand or peak times.

Smart Parking Management System makes the parking process more convenient and efficient, contributing to more efficient use of urban space, reducing traffic congestion, and increasing customer satisfaction. The integration of technology and data analytics plays a key role in transforming traditional parking lots into smart, user-friendly spaces.

Smart Pavement Management System: Smart pavement management systems are technology-driven solutions aimed at monitoring, assessing and managing the condition of pavements and infrastructure. These systems use a variety of technologies and data analytics to enable efficient maintenance, improve road safety, and extend the lifespan of transportation infrastructure. The key components and features of a smart pavement management system are:

- **Pavement Condition Monitoring:** Sensors and imaging technology, such as laser scanners and cameras, are used to assess road surface conditions. This includes detecting cracks, potholes, ruts, and other signs of damage.
- **Data Collection and Analysis:** Collection of data on road conditions, traffic loads, weather conditions and other relevant factors.
- **IoT Sensors and Smart Infrastructure:** Integrate IoT sensors into pavement or road infrastructure to monitor conditions in real time. These sensors can provide continuous data on temperature, humidity, and other factors that affect pavement health.
- **Machine Learning and Predictive Analytics:** Implement machine learning and predictive analytics algorithms to predict future conditions of pavements based on historical data. This enables proactive maintenance planning and resource allocation.
- **Automated Inspection and Evaluation:** Introduce automated systems for pavement inspection to reduce the need for manual inspection. Utilizing drones equipped with cameras and sensors enables efficient and comprehensive evaluation.
- **Dynamic Maintenance Scheduling:** Develop algorithms to dynamically plan maintenance activities based on real-time conditions, budget constraints, and priorities. This ensures optimal use of resources and minimal disruption to traffic flow.

- **Life Cycle Cost Analysis:** To find out economic viability of various maintenance strategies life cycle cost analysis is performed. This also takes into account the cost of renovations, rebuilding, and ongoing maintenance.
- **Remote Monitoring and Control:** Allows remote monitoring of road conditions and control of maintenance processes. This allows authorities to make real-time decisions and respond quickly to emerging issues.
- **Environmental Aspects:** Consideration of environmental factors such as climate, soil conditions and pollution when assessing road conditions.
- **Public Participation and Reporting:** Public participation via mobile app or his website to report road problems. Providing real-time updates on maintenance activities and progress helps build transparency and trust in the community.

Intelligent pavement management systems play a key role in ensuring the durability, safety and sustainability of road infrastructure. Using technology and a data-driven approach, transportation authorities can increase pavement resiliency, reduce maintenance costs, and improve the overall performance of road networks.