

# 1. INTRODUCTION

Data mining has played a significant role in addressing various pattern recognition challenges and has brought significant advancements in research. Data mining is the process of extracting useful information from large data sets. It involves the use of advanced statistical techniques, machine learning algorithms, and database management tools to discover patterns, relationships, and insights in data that are used for decision-making and strategic planning. Data mining is a critical component of modern data-driven businesses and organizations, as it allows them to make sense of the vast amounts of data collected and use it to improve their operations, gain a competitive advantage, and drive growth. Common applications of data mining include customer segmentation, fraud detection, predictive maintenance, marketing campaign optimization and time-series predictions.

Time series data analysis is the process of using statistical techniques and mathematical models to analyse and understand trends and patterns in data that are collected over time. Time series data analysis often involves the use of techniques such as time series decomposition, trend analysis, and forecasting. It requires specialized tools and methodologies to handle the temporal aspect of the data, such as traditional time series models, machine learning and deep learning models. The goal of time series analysis is to extract meaningful insights from the data and make predictions about future trends and patterns. Time series-based data mining is widely used in fields such as finance, economics, weather forecasting, and quality prediction, to make predictions and identify hidden patterns in the data.

This study aims to create predictive models for forecasting the water quality index of rivers, which is crucial for various water quality applications. The models are constructed using advanced deep learning architectures such as the temporal fusion transformer and transfer learning technique.

The chapter begins by introducing data mining and providing a brief overview of the concepts and techniques involved. The significance of water quality and its indexing based on the intended usage of the water is presented in the chapter. The research background, motivation, and objectives are thoroughly outlined, providing a comprehensive understanding of the problem statement and goals of this study.

## **1.1 DATA MINING AND TIME SERIES ANALYSIS**

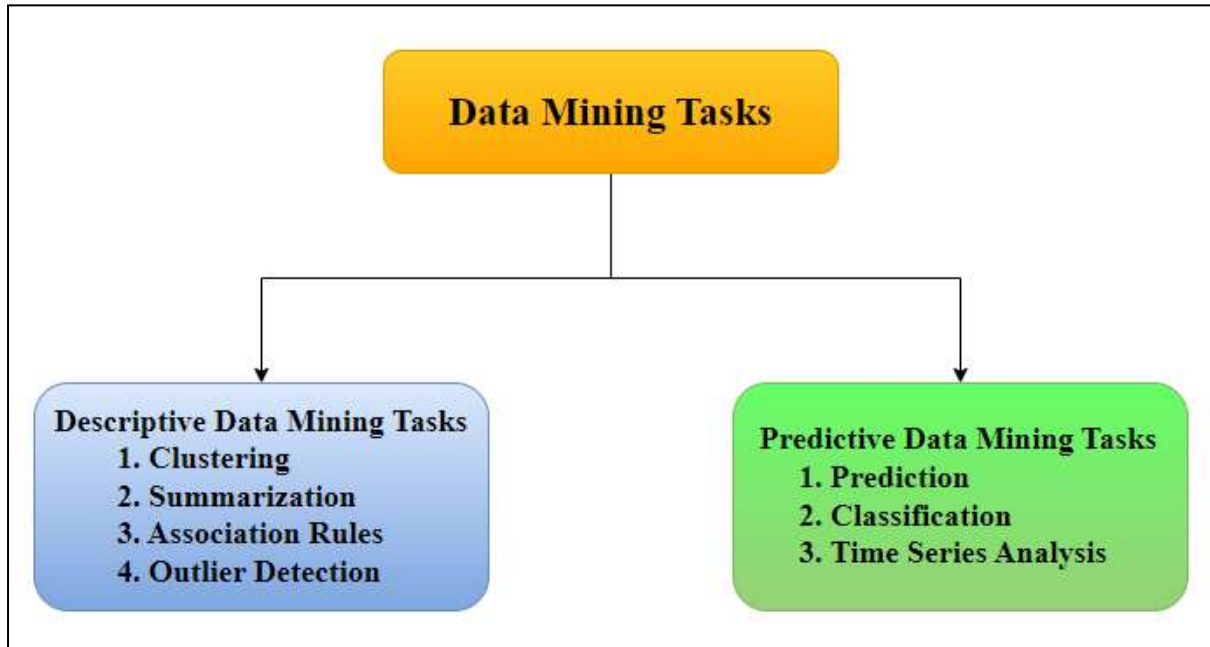
Data mining is a process of identifying patterns in large data sets using techniques from machine learning, artificial intelligence, statistics, and database systems. It is also known as the knowledge discovery process. Various types of data are mined, such as flat files, relational databases, time series databases, data warehouses, transactional databases, multimedia databases, and the World Wide Web. The process of data mining includes gathering requirements, exploring the data, preparing and transforming the data, modelling, evaluating performance, and deploying. The first step is to define the problem and objectives, and then the data is explored and transformed to create metadata. Various tools and algorithms are used for modelling and the results are evaluated, if the results are not satisfactory, the modelling process is repeated. Finally, the results are analysed and deployed in spreadsheets [1].

Data mining techniques, such as classification, clustering, prediction, outlier detection, sequential patterns, and association rules, are crucial for building models and developing applications. Classification is a technique that assigns data points to predefined classes or categories, with the goal of predicting the class for each point in a dataset. Clustering groups similar objects together, while regression analysis examines the relationship between variables. Prediction refers to the process of using data analysis and statistical models to make informed forecasts about future events or outcomes. Accurate prediction requires a thorough understanding of the underlying data and the use of appropriate statistical techniques to extract meaningful insights. Association rule mining uncovers connections between two or more items, while outlier detection identifies observations that deviate from the expected pattern or behaviour. This technique can be used in fields like intrusion detection and fraud detection. Sequential pattern mining finds recurring patterns in data, and prediction combines other data mining techniques to forecast future events based on past events [2]. Data mining is a powerful tool that is being applied in several fields like banking, finance, marketing, telecommunication, healthcare, automobiles, scientific research, text mining, web mining, image mining etc.

### **DATA MINING TASKS**

Data mining tasks refer to the specific objectives or goals that a data mining project aims to achieve. The tasks can be divided into several categories, such as descriptive data mining, predictive data mining, and online data mining [3]. Descriptive data mining tasks focus on summarizing and describing the characteristics of the data, while predictive data mining tasks aim to make predictions about future events or behaviours based on historical data. Online

data mining tasks are focused on analysing data in real-time or near real-time. These tasks can include monitoring and predicting trends in social media data, identifying patterns in financial transactions, or detecting anomalies in sensor data. The selection of the appropriate data mining task will depend on the type and nature of the data being analysed and the specific goals of the project. Fig 1.1 depicts the broad categories of data mining tasks.



**Fig.1.1 Data Mining Tasks**

### **Descriptive Data Mining Tasks**

Descriptive data mining is a type of data mining that is used to summarize, describe, and understand the characteristics of a given dataset. The goal of descriptive data mining is to explore and understand the data, to find patterns, trends, and relationships in the data, and to gain insight into the underlying structure of the data. The distinction between descriptive and predictive tasks lies in their approaches and objectives. The descriptive analysis employs data aggregation and data mining techniques to provide insights and summaries of historical data, aiming to understand past patterns and trends. The predictive analysis utilizes statistical analysis and forecasting techniques to make future predictions. By analysing historical data and applying various algorithms, the predictive analysis aims to anticipate future outcomes or trends. The descriptive analysis focuses on describing and summarizing past data, while predictive analysis leverages statistical analysis and forecasting techniques to make informed predictions about the future [4]. Some of the descriptive data mining tasks such as clustering, summarization, association rule mining and outlier detection are described below.

## ***Clustering***

Clustering is a popular technique in descriptive data mining that involves grouping similar data points into clusters. The goal of clustering is to find patterns and structure in the data by dividing it into meaningful groups based on their similarity. Clustering algorithms use various similarity measures, such as Euclidean distance, to evaluate the similarity between data points and to assign them to the appropriate cluster. The results of clustering analysis provide valuable insights into the structure and relationships of the data and are used for various applications such as customer segmentation, image classification, and anomaly detection. Clustering algorithms are further categorized into hierarchical, density-based, and centroid-based clustering techniques, each with its own strengths and weaknesses. The choice of clustering algorithm depends on the characteristics of the data and the goals of the analysis. There exist various types of clustering algorithms, including partitioning methods, density-based clustering, hierarchical clustering, and k-means clustering, among others [5].

Hierarchical clustering constructs a hierarchy of clusters. Initially, each data point is assigned to its own cluster, and the two nearest clusters are successively merged into the same cluster. The algorithm terminates when only a single cluster remains. The results of hierarchical clustering can be visualized using a dendrogram. K-means clustering is an iterative algorithm that aims to find local minima in each iteration. In this approach, a random number is assigned to the value of k, and each data point is assigned to a cluster. Cluster centroids are computed, and data points are then reassigned to the closest centroid. This process is repeated until local minima are achieved. Clustering finds applications in various domains such as medical imaging, social network analysis, image segmentation, anomaly detection, and more [6].

## ***Summarization***

Summarization is a key aspect of descriptive data mining that involves reducing the size and complexity of the data while preserving its essential information and structure. The goal of summarization is to provide a concise and meaningful representation of the data that are easily understood and analysed. Summarization techniques are either statistical or visualization-based, and include methods such as mean and median calculation, histogram construction, and scatter plot visualization [7]. These techniques allow data analysts to identify patterns, trends, and relationships in the data, and to make informed decisions based on the summarized information. Summarization is especially useful in big data analytics where the sheer size of the data makes it difficult to analyse and understand without reducing its

complexity. The choice of summarization technique depends on the characteristics of the data and the goals of the analysis.

### ***Association Rule Mining***

Association Rule Mining is a technique in descriptive data mining that involves discovering relationships and patterns between variables in large datasets. The goal of association rule mining is to identify correlations and dependencies between variables and uncover underlying rules or regularities in the data. Association rules are generated by analysing transaction data, such as shopping baskets or customer purchase history, to identify frequently occurring combinations of items or events. These rules are then used to make predictions about future behaviour, such as which items are likely to be purchased together. Various algorithms are employed in association rule mining, including Apriori, Apriori Hybrid, and FP-growth [8]. Association rule mining algorithms use measures such as support, confidence, and lift to evaluate the strength of the relationships and to prioritize the most important rules. The results of association rule mining are used in various applications such as market basket analysis, customer segmentation, and recommendation systems.

### ***Outlier Detection***

Outlier detection is a technique in descriptive data mining that involves identifying data points that deviate significantly from the norm. The goal of anomaly detection is to identify outliers or irregular patterns in the data that may indicate a problem or signal the presence of important information. Anomaly detection is used in various applications such as fraud detection, network security, and quality control. Anomaly detection algorithms are either statistical or machine learning-based and use various methods such as standard deviation, clustering, and density-based methods to identify anomalous data points. The results of anomaly detection help organizations detect and prevent fraud, improve product quality, and detect potential security threats. The choice of anomaly detection method depends on the characteristics of the data and the goals of the analysis [9].

### **Predictive Data Mining Tasks**

Predictive analysis in data mining uses mathematical algorithms and machine learning to forecast data, providing valuable insights for business intelligence and research sectors to make better decisions. The process involves training a model with sample data and fine-tuning it to achieve better predictions [10]. The trained model then analyses unknown data to determine its behaviour. Predictive analysis is applied in various business sectors, allowing

companies to gain a competitive edge and predict trends. This approach is categorized into classification, prediction, and time series analysis.

### ***Classification***

Classification is a data mining technique that involves categorizing data into predefined classes based on their attributes or features. The goal of classification is to build a model that can accurately predict the class of new, unseen data based on its attributes. This technique is commonly used in various industries such as finance, healthcare, marketing, and customer service. Various classification algorithms are decision trees, logistic regression, support vector machines, and k-nearest neighbours [11]. Choosing the appropriate algorithm depends on the nature of the data, the size of the dataset, and the accuracy required for the classification task. Classification is a powerful technique that helps organizations to automate decision-making processes and improve their operational efficiency. An attribute set comprises data stored in a specific format, which must be pertinent to the problems under consideration. The classification method encompasses algorithms that require the data to undergo certain procedures, such as data cleaning, relevant data analysis, and data transformation, before being passed to them. During the data cleaning process, noise is eliminated, and missing values are filled using noise removal and filtering techniques. Subsequently, the data is classified using machine learning algorithms, and the evaluation results are obtained along with the class label [12].

### ***Prediction***

Prediction is a data mining technique that involves using statistical models and machine learning algorithms to make predictions about future events or outcomes. This technique is used in various fields, including finance, healthcare, and weather forecasting, among others. The process involves training a model with historical data and using it to make predictions about future events. Predictions can be made for both continuous and categorical variables, and the accuracy of the prediction depends on the quality and quantity of the historical data used to train the model. Several algorithms employed in data mining for prediction include Naïve Bayes, Support Vector Machine (SVM), k-nearest neighbour, decision tree, and artificial neural network, among others [13]. The distinction between regression and prediction lies in their respective purposes. Regression is a technique utilized to discern the relationship between variables. On the other hand, prediction involves the integration of various data mining techniques to forecast future outcomes based on past data [14]. Prediction is a valuable

technique that helps organizations to make informed decisions and gain a competitive advantage.

### ***Regression***

Regression is a data mining technique utilized to predict numerical or continuous values. It involves establishing the relationship between independent variables and dependent variables in statistical analysis. There are several common types of regression in statistics, including linear regression, logistic regression, polynomial regression, stepwise regression, ridge regression, lasso regression, multiple regression, and multivariate regression. Linear regression is a simple form of regression where there is one predictor variable, and the values can be mapped in a two-dimensional space. The predictor values are plotted along the x-axis, while the prediction values are plotted along the y-axis. Logistic regression, on the other hand, is employed when the dependent variable is binary, such as 0 or 1, pass or fail, etc. It explains the relationship between one binary dependent variable and one or more nominal variables [15]. Polynomial regression is a type of linear regression where the relationship between the independent and dependent variables follows an nth-degree polynomial. Stepwise regression is a method for selecting predictive variables in regression models through an automatic procedure. Lasso regression, on the other hand, incorporates shrinkage, where data values are shrunk towards a central point like the mean [16]. In business and financial sectors, linear and logistic regression are commonly used regression techniques. Linear regression is suitable for predicting continuous outcomes, while logistic regression is used for binary outcomes.

### ***Time Series Analysis***

Time series analysis is a statistical technique used to analyse and model time-dependent data. It is an essential tool in many scientific fields, including finance, economics, environmental science, and meteorology, among others. Time series analysis involves examining historical data to uncover patterns and relationships over time, which can be used to predict future trends and make informed decisions [17]. The analysis typically involves identifying and removing trends, seasonality, and other periodic variations to isolate the underlying behaviour of the time series. Time series models can be based on various statistical techniques, including exponential smoothing, Autoregressive Integrated Moving Average (ARIMA), and state-space models. Time series analysis provides valuable insights for understanding complex systems and making predictions based on historical data. It is an essential tool for researchers and practitioners in various fields, contributing to scientific discoveries and advancements.

## **DATA MINING APPLICATIONS**

Data mining applications are used in many industries and disciplines to extract valuable insights from large datasets. These applications involve using machine learning algorithms and statistical techniques to analyse and interpret data to uncover hidden patterns and relationships. In business, data mining applications are used for customer segmentation, market basket analysis, and fraud detection, among others. In healthcare, these applications are used to predict disease outbreaks and diagnose diseases based on patient symptoms. In finance, data mining is used to detect credit card fraud and predict stock prices. Other applications of data mining include social media analysis, text mining, and image recognition. With the increasing availability of large datasets and the development of sophisticated algorithms, data mining is becoming an essential tool for businesses and organizations to gain a competitive edge and make informed decisions. Data mining can be applied to a wide range of fields, including finance, healthcare, marketing, and retail [18]. Some common applications of data mining include:

**Fraud Detection:** Fraud detection is an important application of data mining in the financial industry. Fraudulent activities such as credit card fraud, money laundering, and identity theft can cause significant financial losses to individuals and organizations. Data mining techniques are used to identify and prevent fraudulent activities by analysing large datasets to uncover patterns and anomalies that indicate fraudulent behaviour. These techniques include anomaly detection, predictive modelling, and pattern recognition. For instance, anomaly detection is used to identify unusual patterns of transactions, while predictive modelling is used to predict the likelihood of fraudulent behaviour based on historical data. Fraud detection using data mining is an essential tool for financial institutions to protect their customers and prevent financial losses due to fraudulent activities.

**Customer Segmentation:** Customer segmentation is a data mining application used in marketing to divide customers into groups based on common characteristics such as demographics, behaviour, and purchasing patterns. This technique helps companies to understand their customers better and to develop targeted marketing strategies that are more likely to be effective. By identifying different customer segments, companies can tailor their products and services to meet the specific needs and preferences of each segment. This can lead to increased customer satisfaction, loyalty, and revenue. Data mining algorithms such as clustering and decision trees are commonly used in customer segmentation analysis. Clustering algorithms group customers with similar characteristics together, while decision trees identify



which attributes are most important in determining customer behaviour. Customer segmentation using data mining is an essential tool for companies to gain a competitive advantage and to stay ahead of their competitors in today's fast-paced business environment.

**Predictive Modelling:** Predictive modelling is a data mining technique used to predict future outcomes based on historical data. It involves the use of statistical algorithms and machine learning techniques to analyse large datasets and identify patterns and relationships between variables. Predictive modelling can be used in a wide range of applications, such as weather forecasting, fraud detection, and customer behaviour prediction. By analysing historical data and identifying patterns, predictive models can be trained to make accurate predictions about future events. The accuracy of predictive models can be improved by using more data and refining the model parameters. Predictive modelling is an essential tool for businesses to gain valuable insights into their operations, make informed decisions, and stay ahead of their competitors.

**Market Basket Analysis:** Market basket analysis is a data mining application used in retail to uncover patterns and relationships between products that are frequently purchased together. This technique involves analysing customer transaction data to identify which products are commonly purchased together and to understand the relationships between different product categories. By understanding these relationships, retailers can develop targeted marketing strategies and optimize their product offerings to increase sales and customer satisfaction. Market basket analysis uses techniques such as association rule mining and frequent itemset mining to identify product relationships. Association rule mining identifies which products are frequently purchased together, while frequent itemset mining identifies sets of items that are commonly purchased together. Market basket analysis is an essential tool for retailers to improve their operations, increase revenue, and stay ahead of their competitors in today's competitive retail environment.

**Inventory Management:** Inventory management is a data mining application used to optimize inventory levels and reduce costs in retail and manufacturing industries. It involves analysing historical sales data and forecasting future demand to ensure that the right amount of inventory is available at the right time. By optimizing inventory levels, companies can reduce the costs associated with overstocking or understocking and improve their overall profitability. Data mining techniques such as regression analysis, time series analysis, and decision trees are commonly used in inventory management to forecast demand and make informed decisions

about inventory levels. For instance, regression analysis can be used to identify the relationships between sales and external factors such as promotions and weather conditions, while time series analysis can be used to forecast future demand based on historical data. Inventory management using data mining is an essential tool for businesses to improve their operations, reduce costs, and stay ahead of their competitors in today's competitive market.

**Text Mining:** Text mining is a data mining application used to analyse and extract valuable insights from large volumes of unstructured text data, such as emails, social media posts, customer reviews, and news articles. It involves applying natural language processing and machine learning techniques to identify patterns and relationships between words and phrases in text data. Text mining can be used in various applications, such as sentiment analysis, topic modelling, and entity recognition. For instance, sentiment analysis can be used to identify the sentiment of customer reviews or social media posts, while topic modelling can be used to identify the main topics discussed in a large volume of text data. Text mining is an essential tool for businesses to gain valuable insights into customer feedback, market trends, and competitor activities. By analysing unstructured text data, businesses can make informed decisions and develop targeted marketing strategies to improve customer satisfaction and increase revenue.

**Social Media Analysis:** Social media analysis is a data mining application used to extract insights from social media platforms such as Facebook, Twitter, and Instagram. It involves analysing user-generated content to understand customer sentiment, preferences, and behaviour. Social media analysis can be used to monitor brand reputation, identify customer needs, and develop targeted marketing strategies. Data mining techniques such as sentiment analysis, topic modelling, and network analysis are commonly used in social media analysis to extract valuable insights. Sentiment analysis can be used to identify the sentiment of social media posts related to a brand or product, while topic modelling can be used to identify the main topics discussed in social media conversations. Network analysis can be used to identify the key influencers and communities in social media platforms. Social media analysis is an essential tool for businesses to gain a better understanding of their customers and to develop targeted marketing strategies that are more likely to be effective.

**Image Recognition:** Image recognition is a data mining application used to identify and classify objects and patterns in digital images. It involves analysing visual data and extracting relevant features using computer vision techniques such as neural networks, deep learning, and pattern

recognition. Image recognition has various applications such as object recognition, facial recognition, and image retrieval. For instance, object recognition can be used to identify specific objects in an image, while facial recognition can be used to identify individuals in an image or video. Image retrieval can be used to search for specific images based on their visual characteristics. Image recognition is an essential tool in various fields, including healthcare, security, and entertainment, where it can be used to improve diagnoses, enhance security systems, and create interactive games and experiences.

Data mining requires large amounts of data, powerful computational resources and skilled professionals to extract insights from the data. The process of data mining is iterative, interactive and time-consuming. It also requires careful evaluation of the data, as well as the use of appropriate tools and techniques to extract meaningful insights.

### **TIME SERIES ANALYSIS**

Time series analysis is a statistical technique that is used to analyse time-series data, that is collected over a period of time [19]. The goal of time series analysis is to understand and make predictions about the behaviour of a time-based variable, such as stock prices, weather patterns, or sales data. Time series analysis techniques are either statistical or machine learning-based, and include methods such as trend analysis, seasonality analysis, and forecast models. These techniques allow data analysts to identify patterns and trends in the data, make informed predictions about future behaviour, and make data-driven decisions. The data are used to make predictions about future events or to identify trends or patterns in the data. The basics of time series analysis, include the different types of time series data, the methods used for analysing time series data, and the applications of time series analysis.

Time series analysis is a powerful technique used to analyse data collected over a period of time in various fields such as finance, economics, and engineering. Different types of time series analysis methods can be used depending on the data's nature and analysis goals. Time series analysis techniques are applied in conjunction with other methods such as clustering and association rule mining to gain a complete understanding of the data. The results of time series analysis are used to make informed decisions and optimize processes and systems in various applications such as stock market prediction, sales forecasting, and weather forecasting. In time series analysis, data must be pre-processed, cleaned, and analysed considering dependencies, context, and factors such as outliers, trends, seasonality, and noise. The accuracy of time series

predictions is assessed using metrics such as mean absolute error and mean squared error, and the results can be visualized using techniques such as line graphs and time-based heatmaps.

Time series data is a collection of observations or measurements taken at different time intervals. It is classified into two types: univariate and multivariate. Univariate time series data consists of a single variable measured over time, while multivariate time series data consists of multiple variables measured over time. Univariate time series analysis and forecasting can be performed using techniques such as ARIMA models, while multivariate time series analysis and forecasting can be performed using techniques such as vector autoregression models. The choice of technique depends on the nature of the data and the objective of the analysis.

Time series data can also be characterized by several internal structures such as trend, seasonality, stationarity, and autocorrelation. Time series data are analysed using trend analysis, cyclical fluctuation analysis, and seasonal pattern analysis. Moving average, exponential smoothing, and ARIMA are some of the common types of time series models used for analysis. Time series data is used in various fields such as statistics, signal processing, pattern recognition, econometrics, mathematical finance, weather forecasting, earthquake prediction, electroencephalography, control engineering, and astronomy. There are several methods for analysing time series data, including trend analysis, seasonal analysis, and time series forecasting.

Trend analysis is a technique used in technical analysis that attempts to predict future movements of a stock price based on recently observed trend data. Trend analysis is used to identify long-term trends in the data. It uses historical data, such as price movements and trade volume, to forecast the long-term direction of market sentiment. Trend analysis focuses on three typical time horizons: short-term, intermediate-term, and long-term. It is helpful because moving with trends, and not against them, leads to profit for an investor. Trend analysis is based on a variety of data points, including financial statements, economic indicators, and market data, and there are several different methods that can be used to analyse trends, including technical analysis and time series analysis. Time series analysis is a method of analysing a series of data points collected over a period of time, and it can be useful to see how a given variable changes over time. Trend analysis is often used in finance to analyse stock prices, interest rates, and exchange rates. There are several methods for identifying trends, including visual inspection, linear regression, and exponential smoothing. Visual inspection involves plotting the data over time and looking for patterns or trends in the data. Linear

regression involves fitting a straight line to the data to identify the trend. Exponential smoothing involves smoothing the data to identify the trend.

Seasonal analysis is a crucial technique used in time series analysis to identify and analyse patterns that recur regularly, such as daily, weekly, or monthly patterns [20]. There are various methods for seasonal analysis, including seasonal subseries plots, seasonal indices, and seasonal decomposition. The goal of the seasonal analysis is to identify the seasonal patterns present in the data, which can have a significant impact on the analysis and forecasting of the data. Statistical models such as seasonal averages or more complex models like seasonal autoregressive integrated moving average (SARIMA) models can be used to make predictions about future values of the data. The seasonal analysis is widely used in different fields, including finance, economics, engineering, marketing, and healthcare, to analyse and forecast time series data affected by seasonal factors. Overall, seasonal analysis is an essential tool that helps analysts and researchers make better decisions and predictions by accounting for seasonal patterns in the data.

Time series forecasting is a valuable technique used to predict future trends and values in time-series data. Time series forecasting is used to make predictions about future values of the time series data. Various methods are used for time series forecasting, including Moving Averages (MA), Exponential Smoothing (ES), and ARIMA models. The process involves analysing historical data to identify patterns and trends and using statistical models to make predictions about future values. Time series forecasting can be used to predict trends, cyclical fluctuations and seasonal issues, and can be used in fields such as finance, economics, engineering, marketing, and healthcare. The choice of model depends on the nature of the data and the objective of the analysis. Time series forecasting can be used to predict future values of economic variables such as Gross Domestic Product (GDP), inflation, and unemployment rates, as well as stock prices, interest rates, and exchange rates. It can also be used to predict future values of engineering data such as temperature, pressure, and vibration. Overall, time series forecasting is a valuable tool that helps analysts and researchers make better decisions by predicting future trends and values in time-series data.

### **Applications of Time Series Analysis**

Time series analysis has many applications in various fields, including finance, economics, engineering, and environmental science [21]. In finance, time series analysis is used to analyse stock prices and make predictions about future stock prices. In economics, time

series analysis is used to analyse GDP growth rates and make predictions about future GDP growth rates. In engineering, time series analysis is used to analyse machine performance and make predictions about future machine failures. In environmental science, time series analysis is used to analyse weather patterns and make predictions about future weather patterns. Time series analysis has a wide range of applications in various fields. Some of the common applications of time series analysis are:

**Finance:** Time series analysis is used in finance to analyse stock prices, interest rates, exchange rates, and other financial data. It helps to identify trends, patterns, and relationships between variables, and to make predictions about future values of the data.

**Economics:** Time series analysis is used in economics to analyse economic data such as GDP, inflation, and unemployment rates. It helps to identify the long-term trends in the data and to make predictions about future values of the data.

**Engineering:** Time series analysis is used in engineering to analyse data from sensors, such as temperature sensors, pressure sensors, and vibration sensors. It helps to identify patterns and anomalies in the data and to make predictions about future values of the data.

**Marketing:** Time series analysis is used in marketing to analyse sales data, customer behaviour data, and other marketing data. It helps to identify trends and patterns in the data and to make predictions about future values of the data.

**Healthcare:** Time series analysis is used in healthcare to analyse patient data, such as blood pressure, heart rate, and glucose levels. It helps to identify patterns and anomalies in the data and to make predictions about future values of the data.

In addition to these applications, time series analysis is used in data mining, pattern recognition, and machine learning, where it can be used for clustering, classification, query by content, anomaly detection, as well as forecasting. Time series analysis is a powerful tool that is used to extract meaningful statistics and other characteristics of the data, and to make predictions about future values of the data.

### **Methods of Time Series Analysis**

Time series analysis comprises methods for analysing time series data to extract meaningful statistics and other characteristics of the data [22]. There are several methods available for time series analysis, including:

**Moving average:** A method that involves calculating the average of a subset of data points to smooth out fluctuations and identify trends. It is a simple and effective method for identifying trends in time series data.

**Exponential smoothing:** A method that involves assigning weights to past observations in a way that gives more weight to recent observations, and less weight to older observations. It is useful for forecasting time series data with a trend and/or seasonality.

**Autoregressive integrated moving average (ARIMA):** A method that involves modelling the autocorrelation and moving average of a time series to make predictions about future values. It is a powerful method for modelling and forecasting time series data with a trend and/or seasonality.

**Spectral analysis:** A frequency-domain method that involves analysing the frequency components of a time series to identify patterns and trends. It is useful for identifying periodic patterns in time series data.

**Wavelet analysis:** A frequency-domain method that involves analysing the time-frequency components of a time series to identify patterns and trends. It is useful for identifying patterns in time series data that vary in frequency over time.

**Cross-correlation analysis:** A time-domain method that involves analysing the correlation between two-time series to identify patterns and trends. It is useful for identifying the relationship between two time series.

**Auto-correlation analysis:** A time-domain method that involves analysing the correlation between a time series and a lagged version of itself to identify patterns and trends. It is useful for identifying the presence of serial correlation in time series data.

These methods can be used to analyse time series data in various fields such as finance, economics, engineering, marketing, and healthcare. Time series analysis is beneficial and is commonly used for economic forecasting, yield monitoring, and predicting business trends and metrics. With the increasing availability of data and advances in technology, time series analysis is becoming even more relevant and valuable for decision-making and problem-solving.

## **1.2 OVERVIEW OF WATER QUALITY AND WQI PREDICTION**

Water is an essential natural resource that plays a vital role in the life of humans, animals, and plants. The quality of water is a significant determinant of its suitability for different purposes, including domestic, industrial, agricultural, and ecological [23]. Water quality is defined as the physical, chemical, and biological characteristics of water that determine its suitability for specific uses. The quality of water is influenced by both natural and human activities, such as weather patterns, land use practices, pollution sources, and water management practices. Water quality is a critical environmental issue, and its monitoring and management are crucial for maintaining the health of ecosystems and human populations.

Water quality is a critical factor affecting human health and the environment. Poor water quality can lead to the occurrence of waterborne diseases and other health problems. Over the past 50 years, the water quality and drinking water of the nation have been a challenge. Water quality prediction is an essential aspect of water management, and it involves the use of various techniques to predict the quality of water in a given location.

### **WATER POLLUTION**

Water pollution is a grave environmental issue that occurs when harmful substances or pollutants contaminate water bodies, such as rivers, lakes, oceans, and groundwater. This contamination disrupts the natural balance of ecosystems, leading to dire consequences [25]. Industrial discharges, agricultural runoff, improper waste disposal, sewage and wastewater discharge, oil spills, and atmospheric deposition are common sources of water pollution. The presence of pollutants like heavy metals, pesticides, sewage, oil, plastic debris, and pharmaceuticals poses a significant threat to aquatic life and human health. Water pollution causes the death of aquatic organisms, destroys ecosystems, reduces biodiversity, contaminates drinking water sources, and contributes to the spread of waterborne diseases. Effective measures such as strict regulations, improved waste management, wastewater treatment, sustainable agriculture practices, and individual actions are crucial to prevent and controlling water pollution and safeguarding precious water resources.

Water pollution can be categorized into various types based on the sources and nature of the pollutants. Here are some common types of water pollution:

**Surface Water Pollution:** This type of pollution occurs when contaminants enter surface water bodies such as rivers, lakes, and streams. It includes pollutants from industrial discharges, agricultural runoff, sewage and wastewater, oil spills, and litter.



**Groundwater Pollution:** Groundwater pollution refers to the contamination of underground water sources, typically aquifers. It can result from activities such as improper waste disposal, leakage from underground storage tanks, seepage from landfills, and agricultural practices that introduce pesticides and fertilizers into the soil.

**Nutrient Pollution:** Nutrient pollution, also known as eutrophication, happens when excessive amounts of nutrients, primarily nitrogen and phosphorus, enter water bodies. These nutrients come from sources such as agricultural runoff, sewage, and wastewater. It can lead to the overgrowth of algae and other aquatic plants, causing oxygen depletion and harming aquatic life.

**Chemical Pollution:** Chemical pollution involves the introduction of harmful chemicals into water bodies. It includes industrial discharges, improper disposal of hazardous substances, and the use of pesticides and herbicides. Chemical pollutants can have detrimental effects on aquatic organisms and human health.

**Oil Pollution:** Oil pollution occurs when oil or petroleum products are released into water bodies due to oil spills from ships, tankers, pipelines, or offshore drilling activities. Oil pollution has devastating consequences for marine ecosystems, causing harm to marine life, birds, and other organisms.

**Microbial Pollution:** Microbial pollution refers to the presence of harmful microorganisms, such as bacteria, viruses, and parasites, in water sources. It commonly occurs due to the discharge of untreated sewage and wastewater, contaminating drinking water supplies and recreational waters.

**Thermal Pollution:** Thermal pollution happens when there is a significant increase in water temperature, typically caused by the discharge of heated water from industrial processes or power plants. Elevated water temperatures can have detrimental effects on aquatic organisms and disrupt aquatic ecosystems.

**Plastic Pollution:** Plastic pollution occurs when plastic materials, including single-use plastics and microplastics, enter water bodies. It is a growing concern globally and poses risks to marine life, ecosystems, and human health.

Addressing and mitigating these different types of water pollution require comprehensive management strategies, including improved waste management, wastewater

treatment, stricter regulations, sustainable agricultural practices, and public awareness and education.

### **Point Source Pollution**

Point source pollution refers to the contamination of water bodies that are traced back to specific, identifiable sources [26]. These sources release pollutants directly into the environment through distinct discharge points, such as pipes, outfalls, or vents. Point source pollution can include industrial facilities, sewage treatment plants, power plants, and other similar sources. Point source pollution is easier to identify and regulate compared to non-point source pollution because the origin of the pollutants can be pinpointed to a particular source. This allows for targeted monitoring, control, and enforcement of pollution prevention measures. Permits and regulations are often in place to limit the amount and type of pollutants that can be discharged by point sources.

Efforts to control and reduce point source pollution involve implementing various pollution control technologies, treatment systems, and compliance monitoring programs. These measures aim to ensure that point sources adhere to specific pollutant discharge limits and meet environmental standards. Regular inspections and reporting are typically required to verify compliance. Point source pollution can have significant impacts on water quality and ecosystems. It can degrade aquatic habitats, harm fish and other wildlife, and impair the health of human populations dependent on these water bodies. Therefore, effective management and regulation of point source pollution are essential for preserving the integrity of water resources and protecting the environment.

### **Non-Point Source Pollution**

Non-point source pollution refers to the pollution of water bodies that originates from diffuse and widespread sources, rather than from a specific, identifiable point of discharge. Unlike point source pollution, non-point source pollution comes from various activities and land uses that generate runoff or leachate, carrying pollutants into water bodies through surface water or groundwater [27]. Examples of non-point source pollution include:

**Agricultural runoff:** The use of fertilizers, pesticides, and animal waste in agriculture can result in the runoff of excess nutrients such as nitrogen and phosphorus and harmful chemicals into nearby water bodies.

Urban runoff: Stormwater runoff from urban areas carries pollutants like oil, heavy metals, litter, and chemicals from roads, parking lots, and rooftops, which can end up in streams, rivers, and lakes.

Construction site runoff: During construction activities, sediment, chemicals, and debris are washed away by rainwater, contaminating nearby water bodies.

Forestry practices: Improper logging and clearing of land can lead to erosion and sedimentation of water bodies, impacting aquatic ecosystems.

Septic system leakage: Faulty or poorly maintained septic systems can release bacteria, pathogens, and nutrients into the groundwater, potentially contaminating wells and nearby water sources.

Addressing non-point source pollution requires a comprehensive approach and the implementation of Best Management Practices (BMPs). These practices aim to minimize runoff, improve soil conservation, and reduce the use of harmful chemicals. Examples of BMPs include implementing vegetated buffers, adopting conservation tillage techniques, practising responsible pesticide and fertilizer use, and promoting stormwater management practices. Managing non-point source pollution often involves public education and awareness campaigns to encourage individuals, communities, and businesses to adopt environmentally friendly practices. Additionally, government regulations and incentives can play a role in encouraging the adoption of BMPs and reducing non-point source pollution. Due to its diffuse nature and the involvement of multiple sources, controlling and mitigating non-point source pollution requires cooperation and collaboration among various stakeholders, including government agencies, landowners, farmers, and the general public.

Emerging water quality issues are increasingly gaining attention as new challenges arise in the modern world. One of these issues is the presence of microplastics in water bodies. Microplastics, tiny plastic particles, enter water sources through various means, including the breakdown of larger plastic items and the release of microbeads from personal care products [28]. These microplastics pose risks to aquatic organisms and can potentially enter the food chain, affecting human health as well. Another emerging concern is the presence of pharmaceuticals and personal care products in water. These compounds, including prescription drugs and chemicals from personal care products, can enter water bodies through wastewater effluents and runoff. Their potential impacts on aquatic ecosystems and human health are still being studied. Additionally, the rise of emerging contaminants, such as per- and

polyfluoroalkyl substances (PFAS) and endocrine-disrupting chemicals (EDCs), poses significant challenges for water quality management. These chemicals, found in various products and industrial processes, can persist in the environment and have adverse effects on aquatic organisms and human health. Addressing these emerging water quality issues requires a proactive approach, including research, regulation, and the development of innovative technologies to mitigate their presence and protect water resources for future generations.

Water quality can be evaluated based on its physical, chemical, and biological characteristics. Physical characteristics of water include temperature, colour, taste, odour, turbidity, and total dissolved solids. The chemical characteristics of water include pH, dissolved oxygen, nutrients, metals, and organic compounds. The biological characteristics of water include the presence of bacteria, viruses, and other microorganisms. The quality of water can vary widely depending on its source, location, and use.

Water quality is affected by both natural and human activities. Natural factors that impact water quality include rainfall patterns, soil characteristics, geology, and climate. Human activities that affect water quality include industrial discharges, agricultural runoff, sewage treatment, land development, and recreational activities. Polluted water can have serious consequences for the health of humans, animals, and ecosystems. One such technique is the WQI, which is a tool used to assess the overall quality of water based on several parameters.

## **WATER QUALITY PARAMETERS AND STANDARDS**

Water quality standards are guidelines established by government agencies and other organizations that define the minimum acceptable levels of various water quality parameters. These standards are used to assess the suitability of water for various uses, such as drinking, irrigation, and recreation. The standards are based on scientific evidence and consider factors such as human health, aquatic life, and other ecological systems [29]. Water quality standards can include maximum concentrations of contaminants, minimum levels of dissolved oxygen, and other parameters that are relevant to the specific use of the water.

BIS (Bureau of Indian Standards) is the national standards organization of India and it has established water quality standards for various purposes. These standards provide guidelines for the quality of drinking water, water used in industry, and water used for irrigation, among others. The BIS drinking water standards, for example, cover water quality parameters such as pH, total dissolved solids, total hardness, and the presence of contaminants such as iron, lead, and bacteria. The standards also specify the maximum permissible levels of

these parameters to ensure that drinking water is safe for human consumption. BIS also has standards for water used in industrial processes, which are more stringent than drinking water standards due to the potential for higher levels of contaminants. The standards play a key role in ensuring that water in India is safe and suitable for various uses, and they are widely recognized and respected within the country.

These standards define the acceptable levels of various physical, chemical, and biological parameters in water sources, ensuring that it is safe for consumption, recreation, and ecological balance. The water quality parameters with its BIS standard value are illustrated in Table I.

**Table I. Water Quality Parameters and its BIS Standard Values**

<b>Parameters</b>	<b>BIS Standard (Sn)</b>
Temp(oC)	28
pH	8.5
Conductivity	150
Turbidity	5
Total Alkalinity	200
Chloride	250
COD	10
TKN	100
Ammonia	50
Hardness	100
Ca. Hardness	75
Mg. Hardness	30
Sulphate	200
Sodium	200
TSS	300
TDS	1000
FDS	200
Phosphate	0.3
Boron	1
Potassium	2.5
BOD	3
Fluoride	1.5
DO	7.5
Nitrate-N	0.503
TC	100
FC	60

The standards, regulatory bodies and organizations can assess and monitor water quality, identify potential contaminants, and implement appropriate measures to maintain and improve water safety. Compliance with these standards is essential to prevent waterborne diseases, protect aquatic ecosystems, and promote sustainable water management practices for a healthier future.

## **WATER QUALITY INDEX**

The Water Quality Index is a tool used to assess the overall quality of water based on multiple parameters. WQI is a mathematical model that combines various water quality parameters to provide a single value that represents the overall quality of water [30]. The WQI is calculated using a formula that considers different water quality parameters. The parameters used in WQI vary depending on the location, purpose, and availability of data. The WQI formula consists of three main components: a parameter rating system, a weight factor, and a scaling factor [31]. The parameter rating system assigns a rating to each water quality parameter based on its concentration relative to a standard. The weight factor assigns a weight to each parameter based on its relative importance in determining the overall quality of water. The scaling factor is used to convert the weighted ratings into a single score that ranges from 0 to 100.

### **Applications of WQI**

The Water Quality Index is a tool used to assess the overall quality of water based on several parameters [32]. The WQI is widely used in various applications, including:

**Surface water quality assessment:** The WQI is used to evaluate the quality of surface water in rivers, lakes, and other bodies of water. It provides a simple and effective way to communicate water quality information to the public.

**Groundwater quality assessment:** The WQI is used to assess the quality of groundwater in wells and aquifers. It provides a way to monitor the quality of groundwater over time and identify potential sources of contamination.

**Drinking water quality assessment:** The WQI is used to assess the quality of drinking water in public water systems. It provides a way to monitor the quality of drinking water and identify potential health risks.

Environmental impact assessment: The WQI is used to assess the impact of human activities on the environment. It provides a way to monitor the quality of water in areas affected by pollution and identify potential sources of contamination.

Water resource management: The WQI is used to manage water resources and develop strategies to protect and improve water quality. It provides a way to prioritize water quality improvement efforts and allocate resources effectively.

## **WATER QUALITY INDEX PREDICTION**

Water Quality Index prediction is an essential aspect of water quality management, and it involves the use of various techniques to predict the quality of water in a given location. WQI prediction is based on several water quality parameters such as pH, dissolved oxygen, temperature, turbidity, and conductivity. The WQI model is a popular tool for evaluating surface water quality, and it uses aggregation techniques that allow the conversion of extensive water quality data into a single number that expresses the overall water quality at a certain location and time. The objective of WQI prediction is to turn complex water quality data into information that is understandable and usable by the public.

### **Methods of WQI Prediction**

Several methods are commonly used for predicting WQI. These methods involve the measurement and analysis of various water quality parameters to assess the overall quality of water. Here are a few notable methods:

**Empirical Methods:** These methods involve statistical analysis of water quality data collected from various sampling points. Multiple regression analysis or principal component analysis is often used to establish relationships between water quality parameters and WQI.

**Expert Systems:** Expert systems are computer-based programs that use knowledge and rules provided by experts in the field of water quality. These systems utilize databases of water quality information and apply logical reasoning to predict WQI.

**Machine Learning:** Machine learning techniques, such as decision trees, random forests, support vector machines, and artificial neural networks, can be employed to predict WQI. These methods utilize historical water quality data to train models and then use those models to predict WQI for new samples.

**Mathematical Models:** Mathematical models are based on mathematical equations that describe the relationships between water quality parameters and WQI. These models may consider physical, chemical, and biological processes that affect water quality.

It is important to note that the choice of method depends on the availability of data, resources, and the specific requirements of the study or monitoring program. Each method has its advantages and limitations, and the selection is made considering the context and objectives of the water quality assessment.

### **Importance of Water Quality Prediction**

Water quality prediction is a critical task in the prevention and management of water pollution, and it plays a vital role in the development of water conservation projects. The ability to predict water quality is of utmost importance in controlling water pollution and safeguarding the ecosystem. The employment of advanced artificial intelligence and machine learning techniques for modelling and predicting water quality has proven to be highly effective. The prediction of water quality is an indispensable aspect of guaranteeing the safety and sustainability of water resources.

WQI is a numerical representation of the overall quality of water based on various physical, chemical, and biological parameters. By predicting WQI in real-time, scientists, environmentalists, and water management authorities make informed decisions and take timely actions to protect and maintain water quality. Here are some key aspects highlighting the real-time importance of WQI prediction:

**Early Detection of Water Quality Issues:** Real-time WQI prediction allows for the early detection of water quality issues, such as contamination events or pollutant spikes. By continuously monitoring water parameters and predicting WQI, any abnormal changes or degradation in water quality can be identified promptly.

**Source Water Protection:** WQI prediction helps in protecting source water, which is the initial water body from which drinking water is obtained. By monitoring and predicting WQI in real-time, authorities can identify potential threats or pollutant sources near the water sources.

**Efficient Water Treatment and Distribution:** Real-time WQI prediction aids in optimizing water treatment processes and distribution systems. By having accurate and up-to-date information on the water quality, treatment plants can adjust their processes accordingly to meet the required standards.



Environmental Monitoring and Protection: WQI prediction is crucial for monitoring and protecting the health of aquatic ecosystems. Real-time monitoring enables the identification of changes in water quality parameters. By predicting WQI, environmental authorities can take proactive measures to prevent or mitigate the impacts of pollution.

Public Health and Safety: Real-time WQI prediction plays a vital role in safeguarding public health and safety. By continuously monitoring water quality parameters and predicting WQI, authorities can identify potential risks to human health, such as the presence of harmful bacteria, toxins, or excessive levels of contaminants.

Decision Support for Water Management: Real-time WQI prediction provides valuable decision support for water management authorities. By integrating predictive models with real-time data, decision-makers can assess the impact of various management strategies and interventions on water quality.

Real-time WQI prediction is of utmost importance for monitoring, managing, and protecting water resources. It enables early detection of water quality issues, facilitates source water protection, optimizes water treatment processes, safeguards ecosystems and public health, and provides decision support for effective water management. By leveraging predictive modelling and continuous monitoring, can strive for sustainable water resource management and ensure the availability of clean and safe water for future generations.

The advantages of real-time WQI prediction include timely and continuous monitoring, an early warning system, data-driven decision-making, cost and time efficiency, and the integration of IoT and sensor technologies. The limitations of manual calculations of water quality include limited sampling frequency, time delays, spatial variability, limited parameter coverage, labour-intensive and costly processes, and the inability to capture real-time changes.

While manual calculations have their limitations, they still play an important role in validating and calibrating real-time WQI prediction models. The combination of manual calculation and real-time monitoring provides a more comprehensive understanding of water quality dynamics and facilitates effective water resource management.

Through the use of modelling and predictive techniques, researchers can anticipate future water quality values and identify potential sources of pollution. Such insights are of significant importance, as they can inform the development of water conservation projects and assist in the prevention of pollution incidents. Advanced artificial intelligence algorithms and

machine learning methodologies have been successfully employed to predict the water quality index and classify water quality. Such models are effectively generalized and utilized to forecast the water pollution process, providing decision-makers with actionable insights for the prevention of pollution. Taken together, water quality prediction constitutes a fundamental and necessary undertaking in the effective prevention and management of water environment pollution.

### **1.3 NEED FOR THE PROPOSED RESEARCH**

Water quality prediction is a crucial task as it determines the suitability of water for various purposes, such as drinking, irrigation, and industrial processes. Water quality can vary greatly based on several factors such as the presence of pollutants, season, and location. As a result, it is essential to have an accurate model for water quality index prediction. Developing a WQI prediction model using limited data is not sufficient as it leads to inaccurate predictions. This is because the concentration of pollutants in the water varies significantly from day to day. In addition, improper monitoring stations also can produce inconsistent, incorrect and noisy data, making it difficult to develop an efficient WQI prediction model.

The development of predictive models utilizing historical data is a pervasive practice, particularly in the context of forecasting the WQI for aquatic environments. The acquisition of accurate and comprehensive historical data assumes paramount significance in order to ensure the efficacy and reliability of building WQI prediction models. However, developing a model with a limited amount of data can lead to inaccurate predictions because the concentration of pollutants in the water varies day by day. For example, if a model is developed based on data collected during a specific period, it may not account for changes in water quality that occur throughout the year, such as seasonal variations in rainfall or temperature. Therefore, to improve the accuracy of WQI prediction models, it is important to collect and incorporate as much relevant data as possible to capture the variability of water quality over time.

Monitoring stations that are not properly maintained may produce inconsistent or incorrect data due to factors such as equipment malfunction, calibration issues, or human error. This can lead to inaccurate and unreliable data that is not useful in developing an efficient WQI prediction model. Therefore, it is essential to establish and maintain monitoring stations that are properly located and well-maintained to ensure the accuracy and reliability of the data they produce.

The accuracy of water quality data analysis is vital in identifying and addressing any issues in water resources. The limited number of parameters observed at a sampling station is not enough to identify the trends in water quality data. The data collected from sampling stations are not representative of the overall water quality of a particular area, the WQI prediction may be inaccurate. This is because water quality is influenced by multiple factors, and these factors may not be captured by the limited set of observed parameters.

Machine learning algorithms are increasingly being used in a variety of applications, including water quality prediction [33]. Earlier water quality predictions were done with fewer parameters and used traditional machine learning algorithms; the results were not promising. The traditional machine learning algorithms are not well-suited for handling a large volume of data, which is often required for real-time forecasting. This is because traditional algorithms may become computationally expensive and require significant computational resources to process and analyse large datasets. As a result, a better alternative architecture is required to handle large volumes of data and make real-time predictions, such as water quality prediction.

Deep learning has emerged as a powerful tool in predicting WQI due to its ability to automatically learn complex patterns and relationships in large datasets [34]. By employing artificial neural networks with multiple layers, deep learning models can effectively process and analyse vast amounts of water quality data, encompassing diverse parameters and their interdependencies. These models have the capability to capture non-linear relationships and extract valuable features from raw input data, resulting in accurate and robust predictions of WQI. The integration of deep learning in WQI prediction holds great promise for enhancing water management practices, enabling timely interventions, and ensuring the sustainable utilization of water resources for a healthier and more resilient future.

Hence, there is a pressing need for research focusing on the pool of parameters covering physio-chemical, seasonal and spatial parameters, efficient time series data over a period, and advanced learning mechanisms such as Temporal Fusion Transformers (TFT) and transfer learning to build more efficient WQI prediction models. This research helps to ensure that the efficient models are built and the water quality is accurately predicted.

#### **1.4 DEEP LEARNING AND TRANSFER LEARNING**

Deep learning is a subfield of machine learning that utilizes artificial neural networks with representation learning to extract meaningful patterns and relationships from data. Inspired by the intricate structure and functioning of the human brain, deep learning aims to

emulate its capacity for understanding complex information. Through the use of multiple layers, deep learning algorithms progressively uncover higher-level features from raw input data. This hierarchical approach enables the identification of intricate details at lower layers, such as edges in image processing, and the recognition of more abstract concepts, such as digits, letters, or faces, in higher layers. Deep learning has found applications in diverse domains, including computer vision, speech recognition, natural language processing, machine translation, and time series analysis. It serves as a pivotal component of data science, providing valuable tools for data scientists tasked with gathering, analysing, and interpreting vast volumes of data.

## **DEEP LEARNING ARCHITECTURES**

Deep learning models are designed to learn and make predictions based on large amounts of data [35]. The architecture of deep learning models is designed to simulate the structure and function of the human brain, using multiple layers of artificial neurons that are connected to each other and trained using backpropagation.

Deep neural networks (DNNs) are Artificial Neural Networks (ANNs) with multiple hidden layers between the input and output layers. They are a type of deep learning algorithm that uses multiple layers to progressively extract higher-level features from raw input data. DNNs can model complex relationships between inputs and outputs, making them suitable for a wide range of applications, including image processing, speech recognition, natural language processing, and machine translation. DNNs consist of neurons, synapses, weights, biases, and functions, and are typically feedforward networks in which data flows from the input layer to the output layer without looping back.

The process of creating a DNN involves creating a map of virtual neurons and assigning random numerical values, or weights, to them. The DNN then iteratively adjusts these weights to minimize the difference between the predicted output and the actual output, using a process called backpropagation. The weights are updated based on the error between the predicted output and the actual output, and this process is repeated until the error is minimized. The number of hidden layers and the number of neurons in each layer can be adjusted to optimize the performance of the DNN for a specific task.

Deep learning, including DNNs, is a subset of machine learning that imitates the way humans gain certain types of knowledge. It is an important element of data science and is beneficial to data scientists who are tasked with collecting, analysing, and interpreting large

amounts of data. DNNs are particularly useful for supervised learning tasks, such as speech recognition, document classification, and computational biology. While deep learning is similar to machine learning, it differentiates itself through the way it solves problems, as it understands features incrementally, rather than requiring a domain expert to identify the most applied features. A detailed description of some of the deep learning architectures like autoencoders, deep belief networks, convolutional neural networks, recurrent neural networks, long short-term memory, gated recurrent units, temporal fusion transformer, and transfer learning are given below.

### **Autoencoders**

Autoencoder is an unsupervised learning technique used in deep learning to learn efficient data representations by training the network to ignore the noise and reconstruct the input data [36]. Autoencoders consist of two parts: an encoder that compresses the input data into a lower-dimensional representation, and a decoder that reconstructs the original input data from the compressed representation. Autoencoders are used for various applications, including facial recognition, feature detection, anomaly detection, and acquiring the meaning of words<sup>1</sup>. They are also used as generative models that can randomly generate new data that is similar to the input data.

The merits of autoencoders include their ability to learn efficient data representations, their ability to handle large datasets, and their unsupervised nature, which means they can learn from unlabelled data. Autoencoders can also be used for dimensionality reduction, which is useful for reducing the complexity of high-dimensional data. However, autoencoders can suffer from overfitting, which occurs when the model is too complex and performs well on the training data but poorly on the test data. Additionally, autoencoders require a large amount of training data and computational resources, which can be time-consuming and expensive.

Autoencoders have various techniques, including regularized autoencoders, sparse autoencoders, denoising autoencoders, and contractive autoencoders. Regularized autoencoders are effective in learning representations for subsequent classification tasks, while sparse autoencoders are used for feature selection. Denoising autoencoders are used to remove noise from the input data, while contractive autoencoders are used to learn robust features that are insensitive to small variations in the input data. Variational autoencoders are another type of autoencoder that is used as a generative model for generating new data that is similar to the input data.

## **Deep Belief Networks (DBNS)**

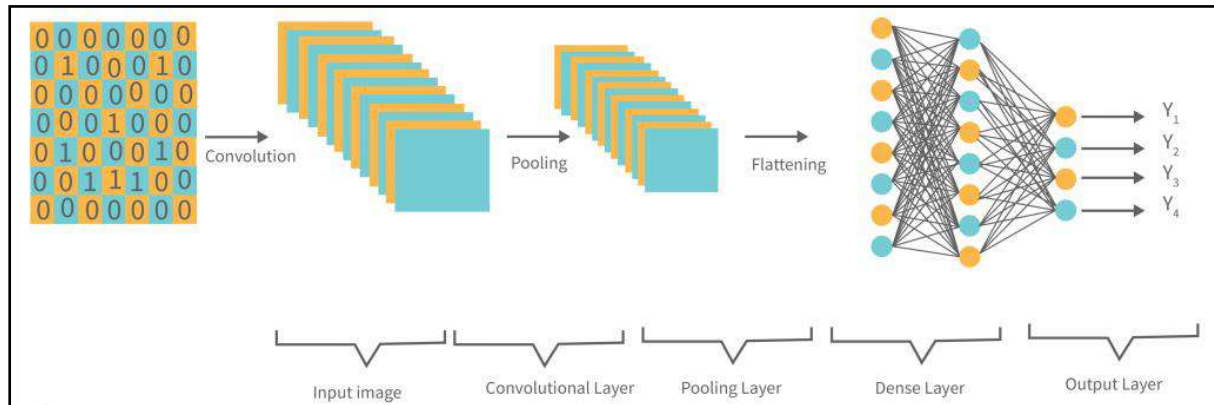
Deep Belief Networks are a type of deep neural network that is composed of multiple layers of Restricted Boltzmann Machines (RBMs) [37]. DBNs are used for unsupervised learning tasks, such as feature extraction, dimensionality reduction, and generative modelling. They can also be used for supervised learning tasks, such as classification and regression. DBNs consist of multiple layers of latent variables, with connections between the layers but not between units within each layer. When trained on a set of examples without supervision, a DBN learns to probabilistically reconstruct its inputs, and the layers then act as feature detectors.

DBNs have various techniques, including greedy layer-wise pretraining, fine-tuning, and dropout regularization<sup>1</sup>. Greedy layer-wise pretraining is used to initialize the weights of the DBN, while fine-tuning is used to optimize the weights of the entire network<sup>1</sup>. Dropout regularization is used to prevent overfitting by randomly dropping out some of the neurons during training<sup>1</sup>. Other techniques used in DBNs include contrastive divergence, persistent contrastive divergence, and deep belief nets with Gaussian visible units. DBNs are generative deep learning algorithms that are used for unsupervised learning. They are composed of multiple layers of RBMs, which are shallow, two-layer neural networks. The first layer of RBM is called the visible layer and represents the input data, while the second layer is called the hidden layer and represents the internal representations learned by the network. DBNs use a layer-by-layer unsupervised pre-training approach, where each RBM is trained one by one, with the outputs of one layer serving as inputs to the next layer. The goal of the pre-training is to learn a hierarchical representation of the input data, such that higher layers represent more abstract features. After pre-training, the DBN are fine-tuned using a supervised learning algorithm, such as backpropagation, to make predictions on a specific task.

## **Convolutional Neural Networks (CNN)**

Convolutional Neural Network is a popular deep learning architecture widely used in image and video analysis, speech recognition and natural language processing tasks. It is designed to automatically learn spatial hierarchies of features from input data. The architecture consists of multiple layers of convolution, activation, pooling and fully connected layers [38]. The convolutional layers are designed to capture the local patterns and relationships between pixels in the input data, and the activation function is used to introduce non-linearity into the network. The pooling layer reduces the spatial resolution of the input, allowing the network to learn robust features. The fully connected layer is used for classification or prediction. CNN

are trained using supervised learning and are able to achieve state-of-the-art performance on many image and speech recognition tasks.



**Fig. 1. 2. Basic Architecture of CNN**

The architecture of a CNN as depicted in Fig.1.2 is purposefully designed for efficient processing of visual data, like images or videos. It comprises key components, including convolutional layers that employ filters to detect specific features in various regions of the input, generating multiple feature maps. Activation functions, like the widely used Rectified Linear Unit (ReLU), introduce non-linearity to enhance the network's ability to model complex relationships. Pooling layers reduce feature map dimensions through operations like max pooling, which selects the maximum value within a region. Fully connected layers at the network's end facilitate high-level reasoning and classification by connecting neurons between layers. The output layer produces task-specific predictions, assigning probabilities to different classes, such as in image classification. CNNs exhibit exceptional performance in computer vision tasks, delivering accurate image recognition, object detection, and image segmentation. The architecture typically follows a sequential structure with layered stacking, accommodating variations based on task complexity and input data size.

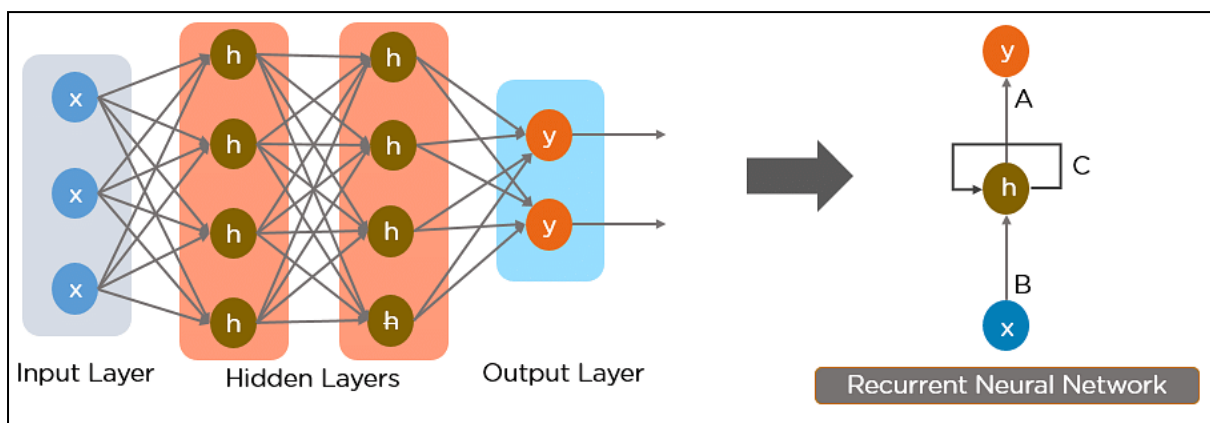
CNNs are widely used in various applications, including image recognition, object detection, and facial recognition. They are also used in healthcare for medical image analysis, disease diagnosis, and drug discovery. In the field of autonomous vehicles, CNNs are used for object detection, lane detection, and path planning. Other applications of CNNs include natural language processing, speech recognition, and recommendation systems.

### **Recurrent Neural Networks (RNN)**

Recurrent Neural Networks are a type of neural network that is specialized for processing sequential data, such as time-series data [39]. RNNs have a unique architecture that

allows them to maintain a memory of previous inputs, making them suitable for tasks such as speech recognition, natural language processing, and image captioning. RNNs consist of a series of interconnected nodes, with each node receiving input from the previous node and outputting to the next node. The output of each node is determined by its current input and its previous output, which allows the network to maintain a memory of previous inputs.

The architecture of RNNs as shown in Fig.1.3, involves processing input sequences one element at a time, with each element being fed into the network as input to the first node. The output of the last node is then used as the output of the network for that input sequence. During training, the weights of the network are adjusted to minimize the difference between the predicted output and the actual output. RNNs use backpropagation through time (BPTT) to update the weights of the network, which involves propagating the error back through time and adjusting the weights accordingly.



**Fig. 1. 3. Basic Architecture of RNN**

The merits of RNNs include their ability to handle input sequences of variable length, their ability to maintain a memory of previous inputs, and their suitability for processing sequential data. RNNs are also effective in learning complex patterns and relationships in data, making them suitable for various applications, including speech recognition, natural language processing, and image captioning. However, RNNs can suffer from the vanishing gradient problem, which occurs when the gradients become too small to be useful for updating the weights of the network.

Additionally, RNNs are computationally expensive, require a large amount of training data and suffer from overfitting, which occurs when the model is too complex and performs well on the training data but poorly on the test data. To overcome the drawbacks of RNNs, various techniques have been developed. One technique is to use Long Short-Term Memory



(LSTM) cells, which are a type of RNN cell that can remember information over long periods of time. LSTMs use gates to control the flow of information, which allows them to selectively remember or forget information from previous time steps. Another technique is to use Gated Recurrent Units (GRUs), which are a simplified version of LSTMs that use fewer parameters and are faster to train. Regularization techniques, such as dropout and weight decay, can also be used to prevent overfitting. Additionally, pretraining the network on a related task can help to improve its performance on the target task.

The drawbacks of RNNs include slow and complex training procedures, the vanishing gradient problem, and overfitting. To overcome these drawbacks, techniques such as LSTMs, GRUs, regularization, and pretraining can be used. These techniques help to improve the performance of RNNs and make them more effective for processing sequential data.

RNNs have various techniques, including LSTM and GRU. LSTM is a type of RNN that is designed to handle the vanishing gradient problem by using a memory cell and three gates to control the flow of information. GRU is a simplified version of LSTM that uses two gates to control the flow of information. Other techniques used in RNNs include bidirectional RNNs, which process input sequences in both forward and backward directions, and attention mechanisms, which allow the network to focus on specific parts of the input sequence.

LSTMs are a variant of RNNs that overcome the vanishing and exploding gradient problems. They have an extra memory component called a cell that allows them to maintain and control the flow of information over a long sequence. GRUs are another variant of RNNs that overcome the vanishing and exploding gradient problems. They have a gating mechanism that allows them to control the flow of information over a long sequence. Gradient clipping is a technique used to prevent gradients from becoming too large. By limiting the magnitude of gradients, this technique helps to prevent the exploding gradient problem.

Overall, RNNs can be challenging to train due to their sequential nature. However, the development of variants such as LSTMs and GRUs, as well as techniques like gradient clipping, have helped to mitigate these challenges and make RNNs a powerful tool for processing sequential data.

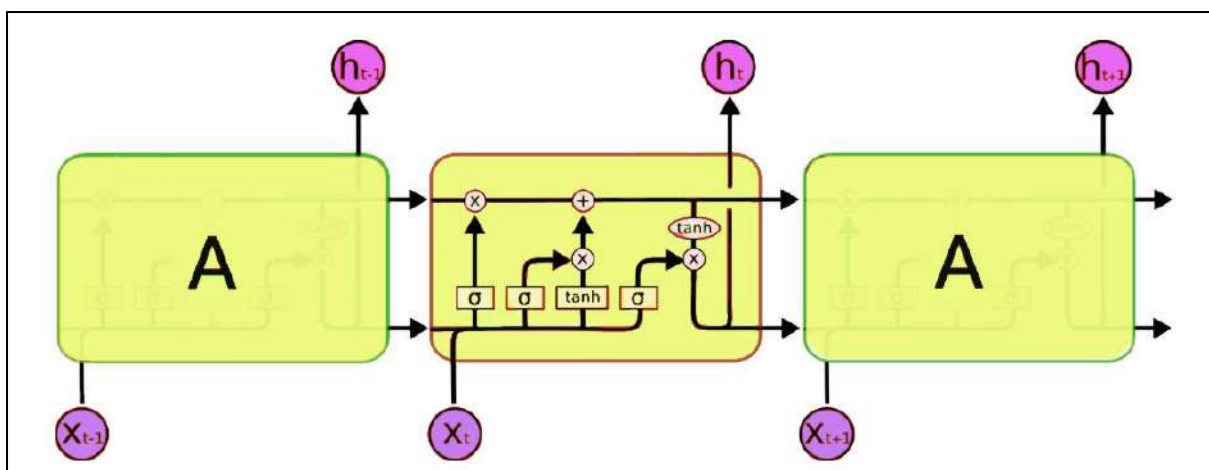
### **Long Short-Term Memory**

Long Short-Term Memory is a type of recurrent neural network that is designed to handle long-term dependencies in sequential data [40]. LSTMs have a unique architecture that allows them to selectively remember or forget information from previous time steps, making

them suitable for various applications, including speech recognition, natural language processing, and image captioning. LSTMs consist of a series of interconnected memory cells, with each cell receiving input from the previous cell and outputting to the next cell. The output of each cell is determined by its current input, its previous output, and its current memory state, which allows the network to selectively remember or forget information from previous time steps.

The working of LSTMs involves processing input sequences one element at a time, with each element being fed into the network as input to the first memory cell. The output of the last memory cell is then used as the output of the network for that input sequence. During training, the weights of the network are adjusted to minimize the difference between the predicted output and the actual output. LSTMs use BPTT to update the weights of the network, which involves propagating the error back through time and adjusting the weights accordingly.

The architecture of LSTM as given in Fig.1.4, has the memory blocks with peephole connections consists of a memory cell, three gates input, forget, output, and peephole connections. The peephole connections allow the gates to access the current value of the memory cell, which helps the gates to make more informed decisions. The input gate controls the flow of new information into the memory cell, the forget gate controls the flow of information out of the memory cell, and the output gate controls the output of the memory cell. The peephole connections allow the gates to use the current value of the memory cell to adjust their behaviour.



**Fig. 1.4 Basic Architecture of LSTM**

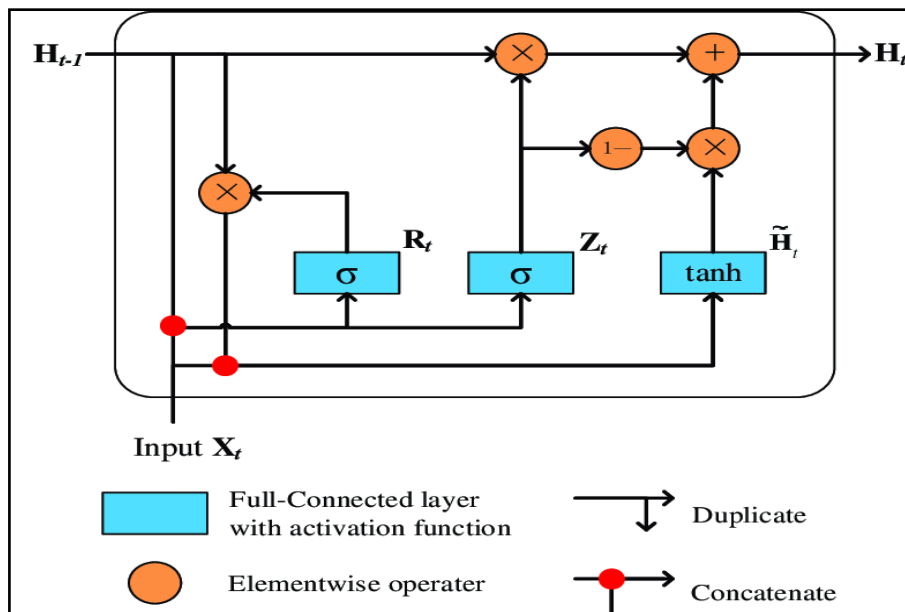
The merits of LSTMs include their ability to handle long-term dependencies in sequential data, their ability to selectively remember or forget information from previous time

steps, and their suitability for processing sequential data. LSTMs are also effective in learning complex patterns and relationships in data, making them suitable for various applications, including speech recognition, natural language processing, and image captioning. However, LSTMs can be computationally expensive and require a large amount of training data. Additionally, LSTMs can suffer from overfitting, which occurs when the model is too complex and performs well on the training data but poorly on the test data. Techniques such as peephole connections, forget gates, input gates, regularization, and pretraining can be used to improve the performance of LSTMs and make them more effective for processing sequential data. LSTMs have various techniques, including peephole connections, forget gates and input gates.

### **Gated Recurrent Unit**

Gated Recurrent Unit is a type of Recurrent Neural Network that was introduced in 2014 as an alternative to the more complex LSTM architecture [41]. GRU is designed to address some of the limitations of basic RNNs, such as the inability to capture long-term dependencies in sequential data and the problem of vanishing gradients. GRU uses gates to control the flow of information through the network, allowing it to selectively remember or forget information from previous time steps. This makes GRU a powerful tool for modelling sequential data, and it has been widely used in applications such as time series prediction, natural language processing, and speech recognition.

GRU consists of a hidden state and a set of gates that control the flow of information through the network [41]. The GRU gates include a reset gate and an update gate. The reset gate controls how much information from previous time steps is used in the current time step, while the update gate determines how much information is passed on to the next time step. During each time step, the GRU takes the current input and the previous hidden state as input and computes a new hidden state based on these inputs and the gate values. The reset gate decides which parts of the previous hidden state to keep and which parts to discard, while the update gate determines which parts of the new hidden state to keep and which parts to update.



**Fig. 1.5 Basic Architecture of GRU Cell**

The GRU architecture allows it to selectively store and retrieve information from previous time steps, enabling it to capture long-term dependencies in sequential data as depicted in Fig.1.5. It also helps to prevent the vanishing gradient problem that can occur in basic RNNs, as the gates allow the network to retain or discard information based on its relevance to the current time step. GRU is a powerful tool for modelling sequential data, as it can capture long-term dependencies and handle vanishing gradients. Its computational efficiency and relatively low number of parameters make it a popular choice for applications where resources are limited.

GRU has several advantages over other RNNs. One of the key advantages of GRU is its ability to capture long-term dependencies in sequential data. This is achieved through the use of gates that allow the network to selectively remember or forget information from previous time steps. The reset gate and update gate in the GRU architecture enable it to selectively store and retrieve relevant information from previous time steps, making it well-suited for tasks such as speech recognition, language modelling, and machine translation. Another advantage of GRU is its computational efficiency. Compared to Long Short-Term Memory, GRU requires fewer parameters and is more computationally efficient. This makes it a popular choice for applications where computational resources are limited, such as on embedded devices or mobile platforms. GRU is a relatively simple and easy-to-implement architecture. It has fewer parameters than other RNNs, which makes it easier to train and optimize. Additionally, the

simple structure of GRU makes it more interpretable, allowing researchers to better understand how the model is making its predictions.

## Temporal Fusion Transformer

The Temporal Fusion Transformer is a cutting-edge deep learning technique for time series analysis that has proven to be effective and efficient across multiple applications. TFT is a multi-horizon model that incorporates a wide range of covariates into predictions and accepts both static and dynamic variables, whose effects are hidden from the user [43]. The model uses a temporal self-attention decoder that enables it to learn long-term patterns by considering the adaptability and architecture of TFT. It is a combination of transformer models, commonly used for natural language processing tasks, and temporal convolutional networks for time series data.

TFT utilizes the transformer architecture and temporal fusion mechanism to construct effective forecasting models for future predictions. The transformer architecture, which is illustrated in Fig.1.6 allows TFT to handle large amounts of data and multiple data sources, such as sensor data and weather data, resulting in a more comprehensive prediction. The temporal fusion mechanism effectively blends temporal information from various data sources to make accurate predictions. This principle of incorporating temporal information from multiple sources enables TFT to account for temporal variability in its predictions.

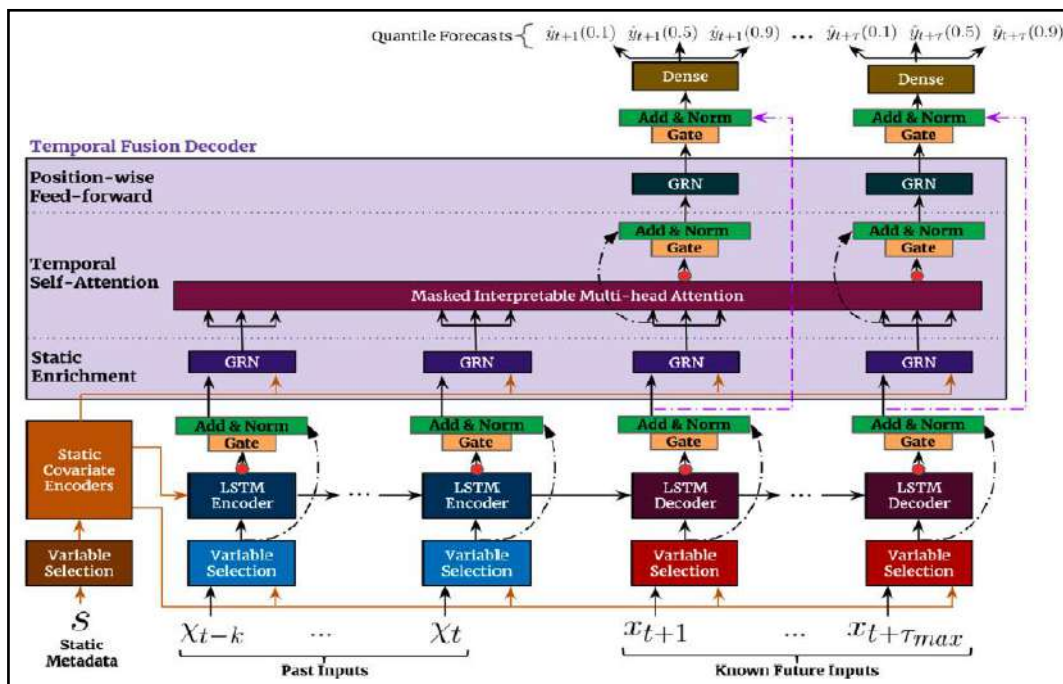


Fig. 1.6 Basic Architecture of TFT

TFT stands out from other deep learning models as it incorporates several key components, including variable selection networks, gating mechanisms, prediction intervals, static covariate encoders, and temporal processing. Variable selection networks allow TFT to select only the most crucial variables for data prediction while gating mechanisms use gated residual networks to selectively produce nonlinear outcomes. Unique static covariate encoders, in the form of Gated Residual Network (GRNs), produce four weight vectors for encoding, which are then conveyed to different TFT layers as needed. TFT also uses a sequence-to-sequence layer for local processing and a multi-head attention block to examine long-term dependencies, allowing the model to learn both short-term and long-term associations between time-varying inputs.

The TFT architecture is a Sequence-to-Sequence model designed for forecasting a univariate target by utilizing static, historical, and future data. The model is composed of gating layers, an LSTM recurrent encoder, an interpretable multi-head attention layer, and a multi-step forecasting decoder. The input size parameter determines the size of the autoregressive inputs, while the static list, historic list, and future list parameters are used to specify the static, historical, and future continuous columns of the data, respectively. The hidden size parameter specifies the number of units in the embeddings and encoders, and the dropout parameter determines the dropout rate for the inputs Variational Sequence Normalization (VSNs). The shared weights parameter can be set to True to enable sharing of parameters between blocks within each stack. The model's activation function can be set to one of several options, including Relu, Soft plus, Tanh, SELU, LeakyReLU, PReLU, or Sigmoid. The loss and validation loss parameters are PyTorch modules instantiated from a collection of train and validation loss functions, respectively. The model's training process is controlled by various parameters, such as maximum steps, learning rate, value check steps, batch sizes, step size, scaler type, random seed, and alias. Finally, the trainer argument parameter allows for the inclusion of additional keyword arguments that are inherited from the PyTorch Lightning trainer. TFT uses the PyTorch Lightning trainer module and Time Series Data Loader to facilitate training and validation.

The LSTM encoder in the TFT converts the input sequence into a fixed-length vector representation, which is then passed to the LSTM decoder for generating the output sequence. The LSTM-decoder layer is improved by combining it with self-attention layers. The Multi-Head Attention (MHA) layer, a key aspect of the Transformer architecture, is a self-attention mechanism that allows the model to focus on different parts of the input simultaneously. In

TFT, the MHA allows the model to focus on different time steps of the input sequence. Composed of multiple sub-layers, each attending to a different aspect of the input, the MHA enables the model to capture both short and long-term trends and learn complex patterns in the input such as temporal correlations. This makes TFT interpretable, allowing it to identify patterns that are immediately obvious.

TFT uses the GRN as a fundamental block. GRN has two activation functions, Exponential Linear Units (ELU) and Gated Convolutional Networks (GLU), and two dense layers. GLU was first introduced to identify the most significant features for prediction. The activation functions help the network determine which input transformations are simple or complex. Output goes through normalization and has a residual connection, potentially bypassing the input. GRN uses static variables depending on its location.

TFT has several advantages over other deep learning models when building prediction models. Compared to RNNs, TFTs can handle sequential data with much longer time steps. TFT is ideal for tasks such as natural language processing and time series forecasting due to its ability to handle sequences of varying lengths and its power in handling sequential data. TFT enhances the interpretability of time series forecasting by identifying globally relevant variables, enduring temporal variations, and significant events for prediction. TFT explains how and why it generates outcomes, unlike black box models, which are difficult to understand. The model supports multiple types of scalers for temporal input normalization and allows the user to specify a custom name for the model. Overall, TFT is a flexible and powerful architecture that can be used in various forecasting applications.

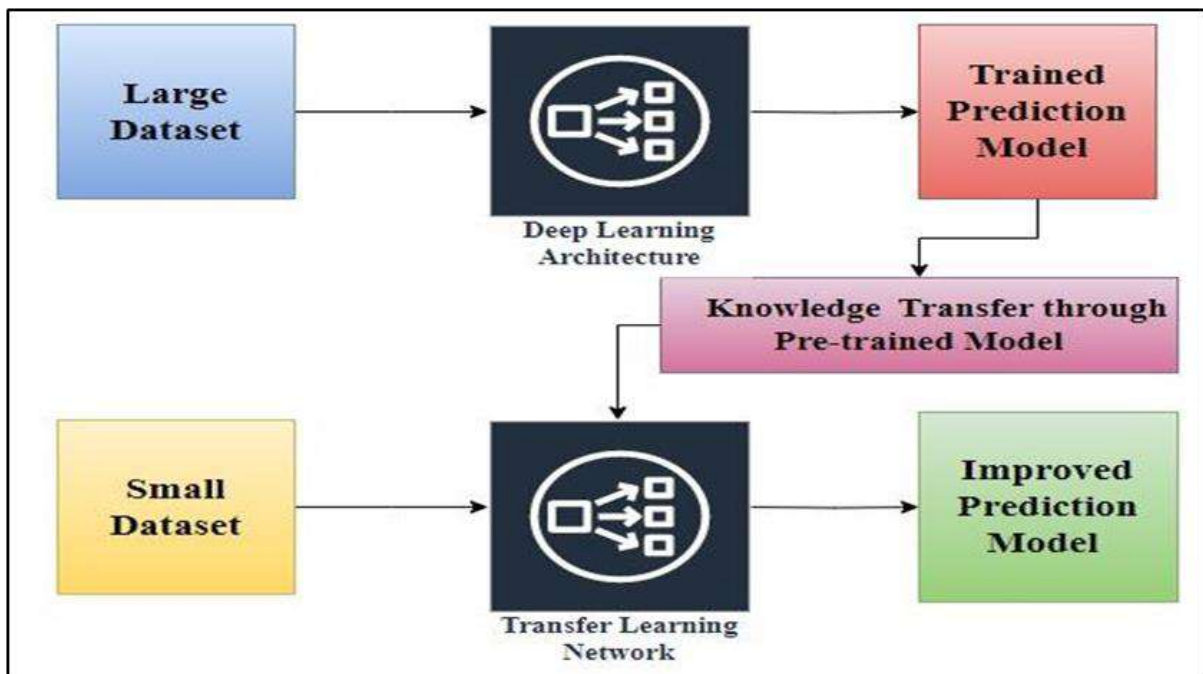
## **TRANSFER LEARNING**

Transfer learning (TL) is a deep learning technique where a model trained on one task is used as the starting point for a model on a second related task. This approach leverages the knowledge learned from the first task and fine-tunes the model to perform the second task, often leading to faster convergence and improved performance [44]. Transfer learning is particularly useful when the second task has limited data available, as the model can utilize the learned representations from the first task to better understand the new data. This technique is commonly used in computer vision and natural language processing, where models trained on large datasets can be adapted to specific use cases with smaller amounts of data.

In transfer learning, there are two approaches to creating a model for a target task: i) developing a model from scratch and (ii) using an existing pre-trained model. The developed

model approach involves training a model specifically for the target task using all available data for that task. This approach is time-consuming and requires large amounts of data to be effective. On the other hand, the pre-trained model approach leverages a model that has already been trained on a similar task to the target task. The pre-trained model is then fine-tuned for the target task, usually by updating its parameters on the target task data. This approach is often faster and requires less data, as the model has already learned features and patterns from the source task that can be applied to the target task. Both approaches have their strengths and weaknesses, and the choice of approach depends on the available resources and the specific requirements of the target task.

Transfer learning has several benefits, including decreased training time, enhanced neural network performance, and the lack of a vast quantity of data. These factors make transfer learning an advantageous approach in many cases. Traditional learning involves training separate models on specific datasets and tasks without retaining any knowledge that can be transferred. In contrast, transfer learning enables leveraging the knowledge that is features, weights, etc. from pre-trained models to train new models, even in situations where data is limited for the new task. Fig. 1.7 illustrates the working principle of transfer learning.



**Fig. 1.7 Working Principle of Transfer Learning**

Transfer learning is a deep learning approach that allows a pre-trained model to be used as a starting point for a new task. This approach allows the model to learn from a larger dataset



than would otherwise be available for the new task, resulting in improved performance. The process of transfer learning typically involves taking a pre-trained model, removing the output layer, and adding a new output layer that is appropriate for the new task. The new output layer is then trained on the new dataset, while the existing layers are frozen or fine-tuned to a lesser degree. By reusing the lower layers of the pre-trained model, the model learns the general features of the data, while the new output layer allows the model to specialize for the new task.

Transfer learning is a machine learning technique that involves using knowledge gained from one task to improve performance on another related task. There are two types of transfer learning: homogeneous and heterogeneous [45].

Homogeneous transfer learning is when the source and target domains have the same feature space, but the marginal probability distributions of the two domains may be different. In this case, transfer can be done through instances, features, or models [46]. Heterogeneous transfer learning is when the source and target domains have differing feature spaces. This type of transfer learning is more challenging than homogeneous transfer learning because it requires bridging the gap between feature spaces and reducing the problem to a homogeneous transfer learning problem [47].

Transfer learning has proven to be a valuable technique in the field of deep learning [48]. It enables the use of knowledge acquired from one task to be transferred to another, similar task, leading to improved performance, reduced training time, and the ability to handle problems with limited data. By combining the strengths of both traditional learning and transfer learning, researchers and practitioners tackle more complex and real-world problems with more efficiency and accuracy. Transfer learning continues to be an active area of research and its applications are rapidly expanding, leading to new breakthroughs and advancements in the field of artificial intelligence.

## **1.5 PROBLEM STATEMENT AND OBJECTIVES OF THE RESEARCH**

The process of accurately predicting the water quality index is a challenging task, primarily due to the fact that water quality is significantly impacted by various environmental factors, including changes in climatic conditions. This research aims to identify the key issues and propose potential solutions to address the issues.

- Developing a model with a limited amount of data is not accurate, as the concentration of pollutants in the water varies day by day.

- Monitoring stations can produce incorrect and inconsistent data which will affect the efficiency of WQI prediction.
- Limited parameters observed in sampling stations will not be sufficient to identify the trends in water quality data, hence the WQI prediction will not be precise.
- Traditional machine learning algorithms cannot handle a large volume of data and the results obtained were not promising hence a more comprehensive learning approach is vital for building a real-time WQI forecasting model.
- River water quality data is a time series-based sequence data that requires specific deep learning architecture to handle temporal patterns.

Hence it is proposed to perform research that will focus on (i) pooled parameters covering physiochemical and seasonal parameters, (ii) suitable time series data over a period and (iii) new learning mechanisms like TFT and transfer learning, to build more efficient WQI prediction models. This research aims to provide a solution to the challenge of accurately predicting river water quality by developing a robust and reliable model. The core objectives of this study are:

- To design and develop an accurate WQI prediction model using physiochemical parameters and RNN and its variants.
- To construct an improved water quality prediction model using pooled parameters and RNN and its variants.
- To build an enhanced WQI prediction model using multi-horizon forecasting with a temporal fusion transformer.
- To create an efficient WQI prediction model using the pre-trained model with RNN variants and homogenous transfer learning.
- To develop a hybrid time series WQI prediction model using temporal fusion transformer and homogenous transfer learning approach.
- To develop a robust time series WQI prediction model using temporal fusion transformer and heterogeneous transfer learning approach.

The proposed approach is unique as it integrates a temporal fusion transformer with transfer learning along with pooled features, which enhances the accuracy and reliability of the prediction.

## **1.6 ORGANIZATION OF THE THESIS**

The rest of the thesis is organized as below. Chapter 2 discusses the literature review of water quality prediction methods using statistical methods, traditional machine learning methods and deep learning approaches.

Chapter 3 presents the problem modelling. This chapter includes data collection, computation of WQI, thorough description of various water quality parameters, Exploratory data analysis, data pre-processing and preparation of datasets. This chapter also explains the methods of model building and performance metrics used for evaluation of the prediction models.

Chapter 4 presents the development of WQI prediction models using deep learning architectures. The chapter first discusses the WQI models developed using physiochemical parameters. The prediction results of WQI models are presented with tables and charts. The performance results of these water quality index prediction models are analysed and compared with the traditional machine learning algorithms.

Chapter 5 presents the development of the WQI prediction model using pooled parameters including physiochemical and seasonal parameters. The importance of adding seasonal parameters to the dataset is explained in detail. Experiments carried out using deep learning architectures such as RNN, LSTM, and GRU are discussed in detail.

Chapter 6 presents the application of temporal fusion transformer architecture in time series analysis. The methodology of building WQI prediction models using the TFT architecture is described. Experiments carried out with temporal fusion architecture are also discussed. The performance results of the water quality index prediction models based on TFT and other deep learning architectures are analysed and reported with tables and charts in this chapter.

Chapter 7 describes the homogenous transfer learning approach for building WQI prediction model in detail. The chapter explains the building of the WQI prediction model using homogenous transfer learning and variants of RNN. The chapter also discusses the building of the WQI prediction model using transfer learning and TFT architecture. The

implementation results of the transfer learning approach are illustrated in tables and charts and the performance analysis is presented.

Chapter 8 deals with creation of the heterogeneous transfer learning based WQI prediction model. The chapter describes the methodology of building the WQI prediction model using heterogeneous transfer learning trained with variants of RNN and TFT. The implementation results of heterogeneous transfer learning are presented in detail and the performance analysis is illustrated.

Chapter 9 summarizes the entire research work carried out in the thesis with research findings. This chapter also presents the achievements of the proposed research work and research contributions. The scope for future research is also stated in this chapter.