**INTRODUCTION**

Modern machine learning models are being used for disease detection and diagnosis, which is causing a revolution in the medical sciences. The use of algorithms to diagnose and offer information on problems like dementia, diabetes, and many types of cancer has shown a lot of promise in recent years, giving hope to the millions of people who experience these conditions.

Many additional critical illnesses, including neuropsychiatric disorders like schizophrenia, may not yet have their secrets unlocked by the machine learning revolution because to the intricacy of these conditions and the paucity of available data. A dangerous and incapacitating mental illness, schizophrenia. There are varying degrees of social function faults, a slow and progressive development of the condition, and manifestations such as impediments to thinking, emotion, and action. False beliefs, clouded or muddled thinking, hearing voices that only you can hear, decreased social engagement and emotional expression, and lack of desire are all symptoms. It not only causes the sufferer enormous suffering, but it also places a significant cost on society and the family. Around 1% of the world's population has the disease. The majority of the datasets for these types of disorders that are now accessible are actually far too tiny to allow for the meaningful extraction of information, and many of these datasets show clear biases about the race, socioeconomic position, and severity of the participants' ailments.

Real-world schizophrenia is a severe neuropsychiatric condition characterized by cognitive symptoms, disorganized thinking, social disengagement, diminished emotional expression, and positive symptoms (hallucinations, delusions). The information we'll look at is made up and mostly made up (poorer working memory and general cognitive ability). Schizophrenia affects 0.5% of people at some point in their lives, and it often first manifests around adolescence and the early years of adulthood.

Schizophrenia can be made better through early detection, intervention, and therapy. The goal of the cure rate is to stop the disease from progressing slowly. The aetiology and pathogenesis of schizophrenia are still unknown, and there aren't any objective laboratory diagnostic markers or equipment diagnostic criteria. Clinical diagnoses are mostly made using a patient's medical history, together with psychiatric symptoms, the disease's law of progression, and scales. The difficulty in making an early diagnosis of schizophrenia is still a result of the intricacy of the disease mechanism. Schizophrenia primarily affects persons between the ages of 15 and 34 and has a severe detrimental impact on human perception, thought, emotion, and conduct. Early controlled, late repeated attacks, and significantly compromised cognitive function are the features of this illness.

Both hereditary and environmental factors raise the risk of developing schizophrenia. Recent research have connected a number of genes to the condition, but it is still unclear how the disorder is impacted by the mix of genetic and environmental factors. The aetiology of schizophrenia is the subject of numerous conflicting ideas, but the fundamental molecular causes of the illness are largely unknown.

The diagnosis of schizophrenia must be made primarily on an individual's behaviour, their medical history, and long-term observations of functional impairment because there are no objective diagnostic tests for the condition. The way this process is carried out varies a lot between countries and even between different medical specialties.

Schizophrenia is especially challenging to identify because of negative signs of the illness such social isolation, symptoms that match with those of other mental illnesses, and cases when sufferers do not immediately exhibit favourable signals. Some of the most frequent psychiatric diagnoses that people obtain initially include bipolar disease, borderline personality disorder, and schizoaffective disorder, which can result in poor treatment management.

With enough high-quality data and a better biological understanding of such conditions, it is hoped that machine learning will one day be able to identify people who exhibit symptoms of schizophrenia more accurately than clinicians, which may aid in the treatment and prognosis of this crippling condition. It is essential that healthcare professionals and machine learning scientists/engineers collaborate in order to fully realise the incredible potential of machine learning in healthcare, while keeping in mind the potentially harmful limitations and side effects that using historical medical data brings limits and drawbacks associated with using previous medical data.

**UNDERSTANDINT THE DATASET**

This dataset contains information on the symptoms and manner of life of 5,000 hypothetical people. The data is contained in a single.csv file with the name "MS4S16 Dataset.csv." The file is comma-delimited and has 5,000 rows and 27 columns (a mixture of numerical and categorical features). If the observation has a value of 1, it refers to a "case," or a person who has the schizophrenia-like illness. If the observation has a value of 0, it refers to a "control," or a person who does not have the schizophrenia-like illness.

**LITERATURE REVIEW**

A synthetic dataset that was presented at the ACM Conference on Fairness, Accountability, and Transparency 2022 served as the basis for the dataset, which has been significantly updated.

The presentation, which was made earlier this year and distributed online, concentrated on the persistent problems with recognised gendered and racialized biases in datasets that degrade model performance and cause serious harm to marginalised people. This is a very crucial part of machine learning that we have just briefly touched upon in the course, despite the fact that it is both a crucial and difficult aspect of the field. This issue needs to be resolved in practically every area of the field that deals with human data, and it will remain a hot topic for many years.