ABSTRACT

College	AMC Engineering College		
Department	Computer Science Engineering		
Student Name	AYUSHI NEGI EKTA KUMARI CHRISTINA S	**USN**	1AM21CS024 1AM21CS054 1AM21CS038
Paper Title	A Supervised Learning Framework for Predicting GSC Antibody Seropositivity in Guillain–Barré Syndrome Using Multivariate Clinical and Demographic Indicators		
Guide	Ms. SANGEETHA RAJ S (Professor Dept. of CSE)		

INTRODUCTION:

★ Guillain-Barré Syndrome (GBS) is an acute immune-mediated polyneuropathy characterized by rapidly progressive weakness, sensory abnormalities, and in severe cases, respiratory failure requiring mechanical ventilation. As the most common cause of acute flaccid paralysis worldwide, GBS represents a significant neurological emergency with an annual incidence of 0.6-2.0 cases per 100,000 population. Despite advances in treatment, approximately 3-5% of patients die from complications, and 20% experience residual disability, underscoring the importance of early and accurate diagnosis .

★ GBS encompasses several clinical variants distinguished by their pathophysiological mechanisms, clinical presentations, and prognosis. These include acute inflammatory demyelinating polyneuropathy (AIDP), acute motor axonal neuropathy (AMAN), acute motor and sensory axonal neuropathy (AMSAN), and Miller Fisher syndrome (MFS), among others. The heterogeneity of these subtypes presents significant challenges for clinicians in diagnosis, treatment selection, and prognostication .

★ Anti-ganglioside antibodies have emerged as crucial biomarkers in the pathogenesis and classification of GBS. Gangliosides are glycosphingolipids highly enriched in the cellular membranes of the nervous system, and antibodies against these structures are frequently detected in GBS patients. Specific antibody-ganglioside associations have been established: anti-GM1 and anti-GD1a antibodies are commonly found in AMAN patients, anti-GQ1b antibodies in MFS patients, and various other correlations between antibody profiles and clinical presentations exist [Kaida et al., 2009]. These antibodies typically develop through molecular mimicry, particularly following infections with pathogens like Campylobacter jejuni, and are detected in up to 50% of GBS patients.

★ Despite their clinical utility, testing for anti-ganglioside antibodies is not universally available, especially in resource-limited settings. Current detection methods such as enzyme-linked immunosorbent assay (ELISA) and immunodot assays require specialized laboratory capabilities, are relatively expensive, and results may not be available during the critical early phase of the disease when treatment decisions must be made [Roggenbuck et al., 2020]. Additionally, the interpretation of antibody profiles can be complex due to cross-reactivity and variability in testing methodologies

★ Recent studies have demonstrated associations between certain clinical features and antibody profiles in GBS, suggesting that it might be possible to predict antibody status based on readily available clinical and demographic information. Machine learning approaches have shown promising results in various medical diagnostic applications but remain underexplored in the context of GBS antibody prediction.

OBJECTIVE:

The primary objective of this study was to develop and validate a supervised learning framework capable of predicting ganglioside complex (GSC) antibody seropositivity in GBS patients using clinical and demographic variables that are typically available during initial patient evaluation. Specifically, we aimed to:

1. Identify the most predictive clinical and demographic features associated with specific antiganglioside antibody profiles in GBS patients.
2. Develop a robust machine learning model capable of accurately predicting the seropositivity of six key anti-ganglioside antibodies (GM1, GM2, GD1a, GD1b, GT1b, and GQ1b) based on routinely collected patient data.
3. . Evaluate the performance of different supervised learning algorithms in this prediction task and determine the optimal approach for clinical application.
4. Assess the potential clinical utility of the predictive framework as a decision support tool, particularly in settings where antibody testing may be delayed or unavailable.

This research builds upon previous immunological studies in GBS, particularly the work highlighting the role of cytokine profiling and multi-lineage T-helper (Th) responses in GBS pathophysiology, as described by Debnath et al. (2018). By integrating machine learning approaches with clinical insights, we aim to bridge the gap between immunological markers and clinical decision-making in GBS management.

METHODOLOGY:

★ Data Collection and Preparation

This study utilized a comprehensive dataset derived from previous research by Debnath et al. (2018) and other published GBS cohort studies. The combined dataset includes records from 129 well-characterized GBS patients with complete clinical, demographic, and serological profiles. Each patient record contains 21 variables: demographic information (gender, age), clinical presentation features (symptom duration, antecedent events), diagnostic classifications (electrophysiological subtype, clinical variant), treatment parameters (mechanical ventilation, hospital stay duration), antibody test results (GM1, GM2, GD1a, GD1b, GT1b, GQ1b positivity), and outcome measures (disability sum score, Erasmus outcome score, neuropathy symptom score, MRC sum score change, Hughes disability change).

In accordance with established research protocols from the referenced studies, GBS patients were diagnosed based on standard clinical criteria and underwent extensive neurological

examinations, electrophysiological studies, and serological testing. Anti-ganglioside antibody status was determined using validated ELISA or immunodot assays in the original studies, with standardized cut-off values for positivity.

Data integration across multiple source studies required careful harmonization of variables and values to ensure consistency. We implemented a systematic approach to resolve discrepancies in clinical terminology, measurement scales, and antibody testing methodologies across the source datasets. Where necessary, we consulted with clinical experts to establish equivalence between different classification systems used in the original studies.

★ Feature Selection and Engineering

Our feature selection methodology was informed by both statistical analysis and domain knowledge from established GBS literature. We utilized a hybrid approach combining:

1. Filter methods: Chi-squared tests, information gain, and correlation-based feature selection to identify variables with statistically significant associations with antibody status.
2. Wrapper methods: Forward selection, backward elimination, and recursive feature elimination with cross-validation to identify optimal feature subsets for each antibody type prediction.
3. Clinical relevance assessment: Features were also evaluated based on their established associations with GBS subtypes in the literature, with particular attention to relationships documented in studies by Kaida et al. (2009), Kim et al. (2014), and The GBS Classification Group (2014).

Previous research has established connections between certain clinical presentations and antibody profiles. For example, ophthalmoplegia and ataxia are associated with anti-GQ1b antibodies, while pure motor symptoms without sensory involvement often correlate with antiGM1 and anti-GD1a antibodies. We incorporated these known relationships into our feature selection process, creating composite variables when appropriate to capture these clinical patterns.

We also engineered features based on temporal relationships, such as the interval between antecedent infection and symptom onset, as these have been suggested to influence antibody development in GBS.

★ Model Development and Evaluation

Building on the methodological framework described by Alarcón-Narváez et al. (2021) for GBS subtype classification, we implemented and compared four supervised learning algorithms:

1. Support Vector Machines (SVM) with different kernel functions (linear, polynomial, radial basis function) to capture complex relationships between clinical features and antibody status.
2. Random Forest (RF), leveraging its ensemble approach to handle the relatively small dataset size while maintaining robust performance.
3. Decision Trees (C4.5 algorithm), selected for their interpretability and ability to model decision processes similar to clinical reasoning.
4. k-Nearest Neighbors (k-NN), which has demonstrated effectiveness in previous GBS classification studies.

For each antibody type, we developed separate binary classification models to predict seropositivity. Additionally, we explored multi-label classification approaches to account for potential correlations between different antibody types, as some patients may be positive for multiple antibodies simultaneously.

To address the class imbalance inherent in rare antibody types, we implemented appropriate balancing techniques including SMOTE (Synthetic Minority Over-sampling Technique) and class weighting. Model optimization was performed using grid search with 5-fold crossvalidation to identify optimal hyperparameters.

Performance evaluation utilized a stratified 10-fold cross-validation approach with balanced accuracy as the primary metric, supplemented by sensitivity, specificity, positive predictive value, negative predictive value, and area under the ROC curve (AUC-ROC). We also employed the Wilcoxon signed-rank test to compare statistical differences between model performances, following the approach described in previous machine learning studies for GBS classification.