Summary

Time series data are ubiquitous across scientific, industrial, and medical domains, capturing the temporal dynamics of complex systems. Extracting meaningful patterns from these sequences is essential for tasks such as classification, anomaly detection, and forecasting. However, time series analysis often presents unique challenges, such as temporal dependencies, variable lengths, and noise, that require the development of specialized methods. A promising research direction is to extend well-established approaches originally designed for vectorial data to the temporal domain. For instance, techniques like Random Forests have achieved remarkable success in tabular data analysis due to their robustness, interpretability, and ability to handle nonlinear relationships. Yet, their adaptation and systematic exploration in the context of time series data remain relatively underexplored. In this thesis, we advance the state of the art of random forests applied to the time series domain along several directions:

First, we develop robust classification frameworks for intraoperative motor evoked potentials (MEPs), time series signals recorded during neurosurgery. Accurate classification of MEPs is critical for monitoring the integrity of motor pathways and guiding surgical decisions. The proposed models leverage Random Forests and forest-based approaches, which can be effectively trained even with limited datasets, providing reliable predictions in high-stakes medical contexts.

Second, we introduce isolation-based models for the early detection of anomalies in MEPs, extending the Isolation Forest paradigm to identify abnormal patterns before they fully manifest. Isolation Forests are among the most widely used state-of-the-art methods for anomaly detection in vectorial data. However, an effective extension capable of handling time series data is missing. The proposed methods address this gap by enabling timely interventions in clinical settings, helping neurosurgeons receive early warnings.

Third, we propose TSRF-Dist, a novel forest-based distance measure that combines an unsupervised tree-based model with the power of Random Forest–derived distances. Recently, distances computed from Random Forests have been developed for vectorial data and demonstrated strong potential in capturing complex relationships between samples. Motivated by these advances, this thesis presents TSRF-Dist, which transposes this knowledge to the time series domain.

Finally, we present tsdistances, a high-performance Python library with a Rust backend that provides efficient computation of elastic distances for time series datasets on both CPU and GPU architectures. While numerous implementations of time series distance measures exist, many remain limited in terms of scalability, extensibility, or computational efficiency. To overcome these limitations, tsdistances integrates several state-of-the-art optimizations for dynamic programming matrix computation, offering a framework that accelerates experimentation and promotes the adoption of time series distances by making their use feasible even for large-scale datasets.

Through these contributions, this thesis provides tools that are accurate, efficient, and accessible to researchers, bridging methodological innovation with practical applications in time series analysis. The main contributions are detailed in the subsequent chapters, including experimental evaluations, comparisons with state-of-the-art approaches, and discussion of their relevance in both medical and general time series contexts.