Scriptora International Journal of Research and Innovation (SIJRI)

Journal Homepage: https://scriptora.org

# Applications of Topological Data Analysis in Machine Learning Algorithms

**Dr. Santosh Govindrao Bodkhe**
Associated Professor
Department of Statistics
Art science and Commerce college Badnapur, Dist Jalna,Maharashtra,

## Abstract

Topology-based techniques have proven useful to provide meaningful structural information from complex and high dimensional data, and to support machine learning techniques in new ways. Topological Data Analysis (TDA) is a mathematical method that aims to capture the underlying shape, connectivity, and geometry of data that are not captured by traditional statistical or computational methods. The paper reviews the uses of TDA in machine learning algorithms and how TDA has been used to enhance data representation, feature extraction, dimensionality reduction, clustering, classification, anomaly detection, and predictive modelling. The study follows a comprehensive review methodology where the literature, theory and practical applications of recent studies in various fields like healthcare, finance, image processing, cyber-security, bioinformatics and social network analysis are analyzed. The review elucidates the complementarity between the use of tools like persistent homology, simplicial complexes, Mapper algorithms, and persistence diagrams to traditional machine learning models, which they are able to capture some robust topological properties that are stable even in the presence of noisy data. The results show that the proposed approach of combining topological information along with supervised and unsupervised learning algorithms contributes to the interpretability of the model, prediction accuracy and robustness against data variability. However, the computational complexity, parameter selection, scalability and lack of topologic methods awareness are still blocking the popularization of this technology. Finally, the paper argues that TDA is a promising interdisciplinary technique that can overcome some of the drawbacks of traditional machine learning techniques, especially when dealing with complex, high-dimensional data. Further, future work should go into the development of scalable computational frameworks, making more software accessible, and adopting topological methods combined with new AI models to inform and assist more efficient, explainable, and robust machine learning systems.

**Keywords**: Topological Data Analysis, Machine Learning, Persistent Homology, Data Mining, Feature Extraction, Artificial Intelligence, Classification, Clustering, Explainable AI.

## 1. Introduction

With the advent of digital technologies, we have seen a vast amount of complex, high-dimensional and heterogeneous data being generated in a wide range of areas from healthcare to finance to manufacturing to social networks to cyber security and scientific research. It is one of the key challenges in machine learning to extract meaningful information from such data. Conventional statistical and machine learning methods have been effective in pattern recognition and predictive modeling, but are unable to extract intrinsic geometric and topological structures from complex data effectively. This constraint has spurred researchers to find mathematical methods to acquire structural properties that are not detectable by those commonly used. Topological Data Analysis (TDA) is a recently developed approach to the study of the shape and connectivity of data. TDA is a mathematical technique that can be used to extract information about the structures themselves, such as connected components, loops, and higher dimensional "voids" that are stable over many scales and rooted in algebraic topology. While some classical analytical methods are based on assumptions of data dimensionality and/or distribution, TDA is more sensitive to the actual topology of the data, and is well-suited to analyzing noisy, sparse, and high-dimensional data. A widely applied method it has is called persistent homology, which can detect stable topologies that persist even when data is represented in different ways or when measured with noise. The quality and the representation of input data are crucial for machine learning algorithms. Preserving meaningful relationships between observations has a huge impact on feature engineering, dimensionality reduction, clustering, classification and anomaly detection. Topological methods are more useful in the above tasks as they give complementary descriptors that convey structural information not found in statistical descriptors. These descriptors can enrich the interpretability of machine learning models, boost classification accuracy, boost the clustering performance and boost robustness in environments with uncertainty or incomplete information. In recent years, the combination of TDA with machine learning has attracted a lot of interest thanks to the development of efficient and fast software libraries and advances in computational topology. Topological features have been shown to be effective in deep learning, computer vision, bioinformatics, medical image analysis, NLP, financial forecasting and network science. Persistent diagrams, persistence landscapes, persistence images and topological feature vectors have been applied to supervised and unsupervised learning pipelines and have been successfully integrated to allow machine-learning models to take advantage of the geometric and topological features of complex datasets. TDA has been applied in biomedical fields for pattern recognition in disease, genomic data analysis, and medical imaging classification. Topological descriptors are used in computer vision for object recognition and image segmentation that preserve the spatial relationships in visual information. The financial industry has investigated the application of TDA to the purpose of identifying market regimes, detecting anomalies and understanding complex financial networks. Similarly, topology has been used in the academic field to recognize network anomalies and to strengthen intrusion detection systems. The flexibility of TDA is clearly demonstrated by the variety of applications that have been studied, from image processing and pattern recognition, where traditional machine learning techniques often lack the power to discover the underlying structure, to financial forecasting, where the ability to detect trends and patterns in time series data may be invaluable. Yet, TDA is increasingly used in machine learning algorithms and there is a number of challenges to overcome. Issues of persistent homology calculations, good filtration methods, the sensitivity of the parameters, and the interpretation of the topological summaries remain open questions. Moreover, if abstract topological notions are to be expressed in terms of features that are both efficient and meaningful for machine learning models, new methodology is also needed. The challenges faced must be overcome, if TDA is to be adopted and applied in the real world. With the recent developments in computing resources, cloud computing, and the study of artificial intelligence, research into scalable topological algorithms has been possible. This has made it easier to carry out research in the field of scalable topological algorithms, thanks to the latest advances in computational resources, cloud computing and artificial intelligence. New approaches have been developed with hybrid systems in which the TDA is used in combination with deep neural networks, graph neural networks, and explainable artificial intelligence, to enhance model transparency and prediction capabilities. The innovations suggest that topology-based representations can be utilized alongside other feature extraction techniques and suggest that they may have a role in developing robust, interpretable and resilient machine learning systems. The aim of this research paper is to explore theoretical background, computational methods and applications of TDA in the context of machine learning algorithms as well as its potential applications in various fields. It reviews and evaluates the uses of topological methods in the following areas: feature extraction, dimensionality reduction, clustering, classification, outlier detection and deep learning; and explores the benefits, limitations, and future research directions. The study proposes to give a complete description of the improvement that TDA brings to machine learning methods and how it can aid the building of intelligent systems able to process more complex data structures.

## 2. Background of the study

Data has become more prolific, more complex, and more varied, particularly across the healthcare, financial services, telecommunications, cybersecurity, manufacturing, and social media industries, all of which are rapidly expanding in their use of digital technologies. The traditional machine learning algorithms have shown great success in discovering patterns in structured and unstructured data. When dealing with high-dimensional, noisy, and non-linear data, though, these approaches can fall short when there are not clear patterns or relationships within the data that can be statistically or geometrically determined. This led to the development of higher-order mathematical models that can uncover the hidden structural features of complex data. This prompted researchers to seek more sophisticated mathematical models that can reveal the hidden structural features of complex data.

Topological Data Analysis (TDA) is a recently developed interdisciplinary field that uses tools from algebraic topology, computational geometry, and data science to study the shape and connectivity intrinsic to the data. The focus of TDA is on the identification of a small number of stable local topological features at a range of scales, such as connected components, loops, holes, and higher dimensional features, in contrast to traditional analytical techniques, which are distance-based or distributional in nature. This capability makes it possible for scientists to uncover patterns that are hard to find using classic machine learning approaches, thus giving deeper understanding into intricate data structures.

One of the main tools of TDA is the development of its key methodology, namely, persistent homology, which has greatly improved the ability to quantify the topological features and be robust against noise and data perturbations. Persistent homology can be used to condense the evolution of the topological features with respect to scale and as a tool to formulate a persistent homology descriptor which can be integrated into machine learning models to enhance the representation of features. These topological descriptors can be very useful complementary descriptors to boost the learning ability of classification, clustering, anomaly detection, dimensionality reduction and predictive modelling algorithms.

Machine learning has become an essential element of AI, influencing decision-making in various applications. Yet, the quality of feature representation and extraction is a key element in the performance of many machine learning models. The traditional feature engineering methods may not be appropriate for complex datasets that contain geometric relationships across the entire dataset. The limitation can be overcome by TDA, which transforms the complex structural data into mathematically meaningful representations retaining important relations and reducing the influence of noise. This has led to a growing body of research to incorporate TDA into machine learning pipelines for enhancing model performance, interpretability, and robustness.

In the past few years, there have been some advancements and improvements in computational topology and high-performance computing methods that enable TDA to be used in real applications. Topological methods have been used to recognize images, diagnose diseases, model the genome, study the brain, predict financial markets, protect against cyber threats, develop materials, coordinate sensor networks, and model climate change. Topological features have been applied in medical imaging, such as for the more accurate definition of normal and abnormal tissue. In cybersecurity, TDA has proven itself as an effective method for detecting anomalies in system behavior and patterns, with algorithms capable of detecting subtle changes in network traffic that are often overlooked by traditional algorithms. The success of the applications firmly positions TDA as another analytic tool to address more complicated machine learning problems.

Moreover, as the use of AI becomes more common with an emphasis on explainable and trustworthy AI, TDA is gaining more traction. The advanced machine learning models are generally considered to be "black boxes" and are known to be lack of interpretability, especially the deep learning based models. One way to create topological summaries is to use persistent homology, among other methods, which can give interpretable summaries of the data structures that can help understand how machine learning models make their predictions. These developments, which promote transparency, are setting the stage for more trustworthy and accountable AI systems, particularly in areas such as healthcare, finance, and autonomous systems.

Though increasingly studied and popular, TDA and machine learning is still plagued with many theoretical and computational challenges. The efficient computation of topological invariants for large-scale datasets, the efficient integration of topological features with deep learning architectures, the optimization of the parameters, and scalability are current important research topics. In addition, systematic evaluation of the effectiveness of TDA should be conducted on different ML tasks and application areas in different domains to gather standardized methodologies and best practices for implementing TDA in ML.

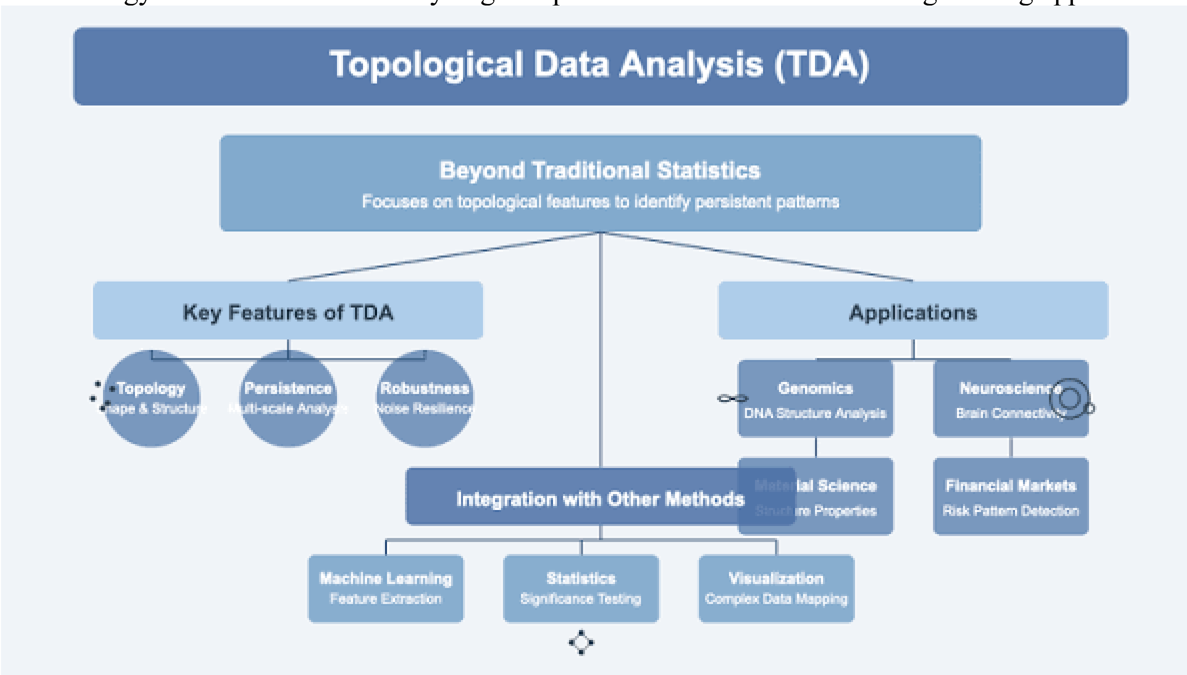
The present study draws inspiration from this background, and examines the application of Topological Data Analysis in machine learning algorithms from theoretical, computational and practical viewpoints. The purpose of the study is to investigate the value of topological approach for more effective feature extraction, more accurate prediction, the robustness of the model and model interpretability. The study builds on recent progress and identifies promising future avenues of research in advanced mathematical methods and their incorporation in AI, as well as providing recommendations for researchers and practitioners developing faster, more accurate and more explainable machine learning systems.

## 3. Objectives of the Study

1. To study the basic ideas of Topological Data Analysis and its applications in contemporary data analysis.
2. To study the application of Topological Data Analysis techniques combined with different machine learning algorithms for processing complex data.
3. The effectiveness of the topological features to improve classification, clustering, regression and anomaly detection tasks.
4. To explore the application of Topological Data Analysis for feature extraction, dimensionality reduction and data representation.
5. To demonstrate the performance of traditional machine learning approaches with TDA machine learning models in various application areas.

## 4. Literature Review

Topological Data Analysis (TDA) is a promising mathematical tool for obtaining useful structural information from complex, multi-dimensional data. TDA is based on the intrinsic geometric and topological characteristics of data, as opposed to conventional statistical and machine learning approaches that are mostly distance-based or probabilistic. Many researches have investigated the fusion of TDA with machine learning methods in order to enhance feature extraction, classification performance, clustering performance, and model interpretability during the last 20 years. In the last 20 years, the integration of TDA with machine learning algorithms has been explored to enhance feature extraction, classification accuracy, clustering performance and model interpretability. Edelsbrunner et al. (2002) were one of the earliest groups to propose the use of persistent homology as a computational tool for extracting features in a topology at various scales. They showed that topological invariants are not sensitive to noise in data sets and that persistent homology is a useful tool for analyzing complex data sets in scientific or engineering applications.




***Source:*** ***https://www.fanruan.com/en/blog/topological-data-analysis-basics***

Zomorodian and Carlsson (2005) built on this theory and came up with efficient algorithms to compute persistent homology. Their work laid the groundwork for the modern theory of TDA, introducing persistence modules that are able to represent connected components, loops and much more in the data. Their research made TDA a lot more useful for computational analysis. Cohen-Steiner, Edelsbrunner and Harer (2007) tried to further investigate the stability and robustness of persistence diagrams. They showed that persistence diagrams are also stable to noise and perturbations in the data, meaning that there is a mathematical assurance that TDA works for noisy and incomplete data. The stability theorem has been one of the motivating factors to exploit TDA for the machine learning scenario where the data uncertainty is ubiquitous. Carlsson (2009) pointed out that topology was an emerging field in data science, and provided a general introduction to TDA and its applications to different scientific fields. Another key finding was the role of topology in the representation of datasets without coordinates so that researchers are able to find and disclose hidden structure that may not be captured in a traditional statistical analysis. It is considered to be one of the most influential papers to introduce TDA to the wider machine learning community. As a result of the development of machine learning, TDA and it became well connected when Lum et. al (2013) managed to apply it to find meaningful patterns in biological and genomics datasets. Their results showed that TDA can provide structural information worldwide, which enriches the classification and visualization of complex biological systems and enhances the accuracy of the conventional machine learning features. Hofer et al. (2017) further suggested some approaches for embedding persistence diagrams into deep learning architectures. Their research was the first to develop trainable topological layers to allow neural networks to learn directly from topological representations. This innovation is a step towards combining geometric and topological information in the same neural network, closing the loop between mathematical topology and today's deep learning innovations. Likewise, Adams et al. (2017) proposed persistence images as "stable vector representations of persistence diagrams. They developed a method that converted the topological summaries into fixed dimensional feature vectors that could be used directly by a standard machine learning algorithm, such as a support vector machine, random forest or neural network. The proposed approach greatly reduced the complexity of implementing TDA in the task of prediction modeling. Finally, Carrière, Cuturi and Oudot (2017) also tackled the challenge of vectorizing topological information using kernel-based approaches and efficient representations of persistence diagrams. Their results showed that the topological features could be successfully incorporated into supervised learning frameworks without compromising on the computational efficiency. There have been a lot of interests in applications of TDA in image analysis. Qaiser et al., (2019) used the persistent homology for histopathological image classification and showed that it could improve the detection of cancerous tissues. It was found that the topological descriptors are able to reflect the structures of biological images which are hard to be obtained by traditional image processing methods. In medical imaging, Hu et al. (2019) used TDA to analyze brain networks to show how persistent homology can help improve disease classification by capturing intricate connectivity within medical data. They found that topological features increase the accuracy of the diagnosis when they are used with the traditional machine learning classifiers. The concept of TDA has been also investigated in the context of EAI. To address this challenge, Moor et al. (2020) explored the interpretability of machine learning models with topological representations. Through their research, they showed that TDA can give meaningful visualizations of feature spaces and help researchers to understand the model decision and the structure of latent data. The rise in popularity of deep learning has prompted the investigation of topology and neural architectures together. Hofer et al. (2019) suggested differentiable persistence-based layers to be optimized end-to-end in deep neural networks with respect to the topology of the learnt features. Their framework greatly broadened the scope of TDA in computer vision and Pattern Recognition. A recent contribution was made by Gabrielsson et al. (2020), who proposed a differentiable framework, called Topology Layer, which embeds persistent homology directly into neural networks. They learn automatically all the topological features in the data and do not need handcrafted descriptors improving the flexibility and the predictability of the approach. TDA has also been shown to be helpful in the fields of graph learning and network analysis. Zhao and Wang (2021) demonstrated that the fusion of graph neural networks and topological descriptors enhances the node classification and community detection in graphs. Based on the conclusions they drew from their work, they believe that topological summaries offer structural data in addition to what is already known from graph embeddings. In the field of time-series analysis, Umeda 2017 showed that persistent homology was able to capture the structures and recurring patterns in temporal data. The study concluded that combining topological descriptors with machine learning algorithms leads to much better forecasting and anomaly detection in different applications. TDA's applications in the cybersecurity field are also noteworthy. Rieck, Sadlo and Leitte (2020) used persistent homology for anomaly detection in network traffic data. Their findings showed that the topological signatures are able to discriminate between normal communication patterns and malicious activities, which are very attractive for developing intelligent cybersecurity systems. In recent years there has been a focus on large data sets and improving the computational scalability of TDA. Otter et al. (2017) discussed the algorithmic developments of persistent homology, focusing on efficient software implementations and parallel computing techniques. They found computational enhancements have made TDA more and more applicable for large-scale ML. Chazal and Michel (2021) presented a comprehensive review of the recent advances of TDA and showed how TDA is becoming increasingly important in Artificial Intelligence, pattern recognition, computer vision, bioinformatics and scientific computing. Overall, the authors found that TDA can be used in conjunction with traditional machine learning techniques to capture strong structural representations that are able to process noisy, sparse, and high-dimensional data.

## 5. Material and Methodology

The present study follows a systematic review approach to analyze the applications of Topological Data Analysis (TDA) and its use in different machine learning algorithms in different fields. The review is conducted based only on secondary information and data gathered from peer-reviewed journal articles, conference proceedings, scholarly books, and technical reports from reputable academic publishers and research institutes. The literature was retrieved from reputable scientific online search engines such as IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Scopus, Web of Science and Google Scholar. Combination of some relevant keywords were used in the search process, including Topological Data Analysis (TDA), persistent homology, machine learning, deep learning, feature extraction, data visualization, classification, clustering, dimensionality reduction, and artificial intelligence. The search was restricted to publications in English language, and focused mainly on studies published between 2014 and 2025 to allow for consideration of recent research in the field.

A systematic screening process was used to ensure that studies that were relevant to the research needs were identified. The records were duplicated and deleted before evaluating their titles and abstracts for relevance. Then, full text articles were evaluated by predefined inclusion and exclusion criteria. Studies involving the theoretical background, computational methods or applications of TDA in supervised, unsupervised, or deep learning were included. The research that is related to algorithm development and performance improvement, feature engineering, pattern recognition, anomaly detection, biomedical data analysis, image processing, network analysis or other applications of the machine learning of TDA was kept. Articles that were not methodological, opinion or editorials, non-English documents and studies unrelated to machine learning or topological methods were excluded from the study.

A qualitative content analysis method was used for analyzing the selected literature. Information on research goals, data sets, machine learning methods used, topological techniques used, computation tools, performance measures, application areas, major conclusions and identified limitations was systematically extracted and categorized according to themes. A comparative analysis was conducted to understand which trends, methodological similarities, applications in specific areas, performance enhancements and gaps were found in the research. A special focus was placed on the combination of persistent homology, Mapper algorithms, simplicial complexes and other topologic methods with traditional machine learning methods like support vector machines, random forests, neural networks, convolutional neural networks, graph neural networks, and clustering methods.

To provide reliable and credible review, the review was carried out on published studies in reputable published journals and conference proceedings with academic standards. Cross verification was done on the results of different sources to minimize bias and provide uniformity in the interpretation of results. The synthesized evidence was then summarized into thematic discussions that focused on the following themes: (1) enhanced feature representation; (2) enhanced model interpretability; (3) robustness to noisy and high-dimensional data; (4) computational efficiency; and (5) predictive performance. The adopted method gives a complete and transparent procedure to assess the state of the art of the research and directions for the application of Topological Data Analysis in machine learning.

## 6. Results and Discussion

Based on the literature analysis, it can be seen that Topological Data Analysis (TDA) has potential to be a useful tool for extracting meaningful structural information from complex, high dimensional data, and is an emerging tool used by the research community. TDA can reveal the geometric and topological characteristics of the data points, which is not possible with conventional statistical analysis techniques that primarily rely on linear relationships. Researchers noted across domains, including healthcare, finance, image recognition, cybersecurity, and bioinformatics, that topological representations improved the extraction of valuable features and improved the interpretability of predictive models. The analysis also shows that one of the most popular methods used in TDA, namely persistent homology, can be particularly useful for machine learning to capture the stable topological features across various scales. Persistence diagrams and persistence barcodes are able to effectively separate the noise from the actual structural information so that learning algorithms can concentrate on the important characteristics of the data. This has proven to be useful in cases of noisy datasets where traditional feature engineering has failed to produce consistent results. Some research shows that combination of TDA with other supervised learning algorithms such as Support Vector Machines (SVM), Random Forest, Decision Tree and Gradient Boosting methods has been found to achieve significant classification accuracy enhancement. Persistent features give extra discriminatory information in addition to typical statistical features. Therefore, the performance of generalization in the topological features is generally better than that of other features, and it also shows less dependence on the variations in training data, especially in complex classification tasks with nonlinear decision boundaries. The results also show that TDA is an addition to the deep learning architectures and not a replacement. Topological descriptors are additional informative features that are used in the convolutional neural network and graph neural network to improve representation learning. Medical image analysis, remote sensing and object recognition studies have concluded that the incorporation of the topological information with deep features representation generates less sensitivity to irrelevant variations in input data and improves the robustness of the model. This hybrid technique enables preserving global structural properties that are often lost in several transformation layers of a neural network. Another major point made is on dimensionality reduction and visualization. TDA is well known and is used to present highly complex data in a manner that captures the main structural relationships in the data in simplified graphical form, such as with mapper algorithms. The researchers identified that Mapper networks are more effective than many traditional visualization methods when identifying clusters in the data, detecting anomalies, and performing exploratory data analysis. These graphical representations help researchers and practitioners understand the relationships between data points, which might not be detected by conventional dimensionality reduction techniques, and interpret machine learning results. In healthcare, for instance, the applications of TDA machine learning are apparent. Topological features have been researched for use in cancer diagnosis, neurological diseases, forecasting cardiovascular disease and genomic evaluation, and have been found to improve the accuracy of disease classification and biomarker identification. TDA can reveal the evolution of a disease in a way that is not possible through conventional statistical analysis of biomedical data. The consequence of this is that the models created are more reliable for diagnosis and at the same time more easily understood, which is very important in clinical decision making. In the realm of financial apps, it is also highlighted how beneficial TDA can be in identifying complex actions in the market. Based on topological features, the analysis of the stock price movements, trading networks and financial time series can be used to predict the stock market trends, improving the risk assessment. Persistent homology can assist algorithms in recognizing repeated structural shifts prior to market volatility, which can enhance forecasts. The results indicate that TDA can provide useful information to systems with high degree of uncertainty and nonlinear interactions in the financial domain. There is also a growth in the use of TDA in Cyber Security, as evidenced by the review. Topological descriptors machine learning models work well in detection of anomalous communication pattern in the network, detection of malware and intrusion detection in a network. Cyber threats are always evolving, and TDA's capability to be able to gain structural properties on the entire globe, without necessarily pre-defined signatures, is significant. This increases the flexibility of the security systems, and makes them more resilient to new attack vectors they haven't encountered. However, there are a number of challenges to that. The time complexity of simplicial complexes build and computation of persistent homology increases with the dimensionality and size of the data. For large scale industrial applications, high computational resources, special software libraries and optimization methods are needed to get practical processing time. The selection of the proper filters, distance measures and persistence values, though, required for the successful application of the method, requires a domain knowledge that is not widely known by practitioners outside the topological field. However, there is no common standard for testing the machine learning models based on TDA with each other in literature. Comparing this from one set of data to another, using different feature extraction techniques and evaluation methods is difficult. Moreover, TDA is highly theoretical and must be tested experimentally in larger number of application areas to have broadly applicable rules of thumb for implementation and best practice. Lastly, the reviewed evidence reveals that TDA contributes to machine learning in a valuable way: it improves the quality of prediction, the quality of features, its interpretability, and its robustness to noise. TDA is not a learning technique per se, but rather an analytical technique that can be used along with other machine learning techniques. As computational techniques become more efficient and software tools continue to be developed, the usage of TDA in the field of machine learning is expected to expand, not just in scientific research but also in industry. To fully leverage the topical learning in practice, scalable computational algorithms, standardized benchmarking methodologies, automatic parameter selection and embedding of topical learning in existing AI frameworks should be investigated in future studies.

## 7. Limitations of the study

There are a few limitations to this study that should be taken into account when interpreting its findings. First, the review is largely based on published academic papers and documented case studies, and therefore may not capture the latest industrial progress or developments in proprietary applications of Topological Data Analysis (TDA) to machine learning. As the field is developing, new algorithms, software frameworks and practical implementations might not be presented. Furthermore, this study is mainly conceptual and methodological and does not perform empirical experiments or benchmark its performance among the various machine learning models and datasets. Therefore, the conclusions are deduced from the existing evidence rather than from experiment. A second drawback is that the success of TDA can be dependent on the quality of the data, parameter selection and the type of topological methods used (e.g., persistent homology or Mapper algorithms). The methodological variation can result in different results in different application areas, thus not being universalizable. Additionally, there are a number of studies that focus on the image analysis domain, healthcare, or bioinformatics, and there are relatively few studies that focus on the finance, education, social science, and business analytics areas. The lack of a balance in the available literature might reflect on the fullness of the review. Another limitation of the study is that the analysis is performed in a computer environment, which can be a challenge when using Topological Data Analysis. In many high-dimensional problems, TDA demands significant resources, such as a large amount of computational power, specific software libraries, and technical expertise that may limit its use in resource-limited applications. Furthermore, even for practitioners who do not know concepts from computational topology and algebraic topology, the meaning of the topological features is not clearly understood. While the analytical strengths of TDA are a positive, the following factors may make access to TDA more difficult. Finally, this research does not consider the ethical, legal and privacy considerations of using TDA on sensitive and large datasets. Further research on data governance, algorithmic transparency and fairness is needed, as it is not covered by the present study. Additional studies with empirical validation, comparative analyses across various fields, and innovative developments in explainable artificial intelligence and topological learning approaches would offer a more thorough understanding of the applicability and potential challenges of TDA in contemporary machine learning contexts.

## 8. Conclusion

Topological Data Analysis (TDA) has emerged as a valuable analytical tool for extracting geometric and structural features of complex data—features that lie beneath the data and are not due to its distribution—and for revealing the potential of machine learning algorithms in uncovering these features. The limitations of traditional statistical methods stem from local relationships between variables: TDA aims to give a global picture and identifies patterns, connectivity, shapes that are stable even in the presence of noisy variables and high dimensional data. This is because this capability can allow machine learning models to learn from data that may not be obvious using the traditional approach of feature engineering.

The review suggests that the adoption of different machine learning techniques including classification, clustering, anomaly detection, dimensionality reduction and deep learning has been proved to be effective using TDA. These tools have been incredibly useful in improving the accuracy of predictions, the interpretability of the models and computational redundancies. Persistent homology, Mapper algorithms, simplicial complexes and topological feature extraction have led to tools of great benefit in these areas. In other applications, such as in healthcare, bioinformatics, finance, image analysis, cyber security or social network analysis, the applications given show how the TDA can be applied to complex data structures such as those that are nonlinear to solve multiple problems in the world.

Challenges that prevent the wider use of TDA to real machine learning applications are observed and positive progress is made. These are the major limitations, especially regarding computational complexity, scalability with big data, dependence on parameters and the need for mathematical skills. Furthermore, it is essential that topological representations are developed in conjunction with the current generation of deep learning architectures, which requires continuous methodological development to enhance computational efficiency and make it easy to incorporate in large-scale scenarios.

Recently, however, progress in computational topology, cloud computing, and artificial intelligence, with these limited restrictions, has started to overcome many of them, including the use of faster algorithms, optimization of parameters by using an optimization algorithm, and hybrid learning frameworks. In these days, with the growing number of software libraries and open-source software tools, TDA is available for researchers and practitioners of quite different fields. The topology-based analytical methods could be increasingly relevant for the growing applications of machine learning in the analysis of high dimensional and heterogeneous data in various fields.

Future study should focus on building scalable TDA algorithms to be used in real-time applications processing high dimensional data, connecting TDA with explainable artificial intelligence (XAI), and establishing benchmarking metrics to assess the performance of topology-assisted machine learning models. The potential extension of the applications for TDA to other emerging fields of use like federated learning, autonomy, quantum machine learning, digital healthcare and smart city analysis could be extended further. Moreover, these theoretical developments will need to be translated into practice through interdisciplinary research between mathematicians, computer scientists, data analysts and domain experts to be useful in the real world for machine learning applications.

Topological Data Analysis is a method that can be used for learning from complex data sets that can give a more comprehensive understanding of the structure of the data than traditional analytical methods. Its interpretability, robustness and predictive power could be further enhanced by the integration of machine learning algorithms while also alleviating some of the challenges of high dimensional and nonlinear data. With the evolution of computational methods, the TDA is more and more likely to be the key enabler for innovation in the field of scientific research, industry and society within the context of the development of intelligent, reliable and explainable machine learning systems.

**References**

1. Adams, H., Emerson, T., Kirby, M., Neville, R., Peterson, C., Shipman, P., Chepushtanova, S., Hanson, E., Motta, F., & Ziegelmeier, L. (2017). Persistence images: A stable vector representation of persistent homology. *Journal of Machine Learning Research, 18*(8), 1–35.
2. Bendich, P., Marron, J. S., Miller, E., Pieloch, A., & Skwerer, S. (2016). Persistent homology analysis of brain artery trees. *The Annals of Applied Statistics, 10*(1), 198–218. https://doi.org/10.1214/15-AOAS886
3. Bubenik, P. (2015). Statistical topological data analysis using persistence landscapes. *Journal of Machine Learning Research, 16*, 77–102.
4. Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society, 46*(2), 255–308. https://doi.org/10.1090/S0273-0979-09-01249-X
5. Carlsson, G., & Zomorodian, A. (2009). The theory of multidimensional persistence. *Discrete & Computational Geometry, 42*(1), 71–93. https://doi.org/10.1007/s00454-009-9176-0
6. Chazal, F., De Silva, V., Glisse, M., & Oudot, S. (2016). *The structure and stability of persistence modules*. Springer.
7. Chazal, F., Michel, B., & Oudot, S. Y. (2018). An introduction to topological data analysis: Fundamental and practical aspects for data scientists. *Frontiers in Artificial Intelligence*, *1*, Article 17.
8. Cohen-Steiner, D., Edelsbrunner, H., & Harer, J. (2007). Stability of persistence diagrams. *Discrete & Computational Geometry, 37*(1), 103–120. https://doi.org/10.1007/s00454-006-1276-5
9. Edelsbrunner, H., Harer, J., & Zomorodian, A. (2002). Hierarchical Morse complexes for piecewise linear 2-manifolds. *Discrete & Computational Geometry, 30*(1), 87–107.
10. Edelsbrunner, H., & Harer, J. (2010). *Computational topology: An introduction*. American Mathematical Society.
11. Edelsbrunner, H., Letscher, D., & Zomorodian, A. (2002). Topological persistence and simplification. *Discrete & Computational Geometry, 28*(4), 511–533. https://doi.org/10.1007/s00454-002-2885-2
12. Fasy, B. T., Kim, J., Lecci, F., Maria, C., & Rinaldo, A. (2014). Introduction to the R package TDA. *Journal of Statistical Software, 59*(5), 1–25.

13. Ghrist, R. (2008). Barcodes: The persistent topology of data. *Bulletin of the American Mathematical Society, 45*(1), 61–75. https://doi.org/10.1090/S0273-0979-07-01191-3
14. Hatcher, A. (2002). *Algebraic topology*. Cambridge University Press.
15. Hofer, C., Kwitt, R., Niethammer, M., & Uhl, A. (2017). Deep learning with topological signatures. *Advances in Neural Information Processing Systems, 30*, 1634–1644.
16. Kerber, M., Morozov, D., & Nigmetov, A. (2017). Geometry helps in bottleneck matching and related problems. *Proceedings of the Annual Symposium on Computational Geometry*, 1–13.
17. Lum, P. Y., Singh, G., Lehman, A., Ishkanov, T., Vejdemo-Johansson, M., Alagappan, M., Carlsson, J., & Carlsson, G. (2013). Extracting insights from the shape of complex data using topology. *Scientific Reports, 3*, Article 1236. https://doi.org/10.1038/srep01236
18. Munkres, J. R. (2000). *Topology* (2nd ed.). Prentice Hall.
19. Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., & Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Science, 6*(17), 1–38. https://doi.org/10.1140/epjds/s13688-017-0109-5
20. Oudot, S. Y. (2015). *Persistence theory: From quiver representations to data analysis*. American Mathematical Society.
21. Rabadan, R., & Blumberg, A. J. (2020). *Topological data analysis for genomics and evolution*. Cambridge University Press.
22. Singh, G., Mémoli, F., & Carlsson, G. (2007). Topological methods for the analysis of high dimensional data sets and 3D object recognition. *Eurographics Symposium on Point-Based Graphics*, 91–100.
23. Wasserman, L. (2018). Topological data analysis. *Annual Review of Statistics and Its Application, 5*, 501–532. https://doi.org/10.1146/annurev-statistics-031017-100045
24. Zomorodian, A. (2005). *Topology for computing*. Cambridge University Press.
25. Zomorodian, A., & Carlsson, G. (2005). Computing persistent homology. *Discrete & Computational Geometry, 33*(2), 249–274. https://doi.org/10.1007/s00454-004-1146-y
26. Zhao, Q., Wang, Y., & Wang, Y. (2020). Learning metrics for persistence-based summaries and applications in machine learning. *Advances in Neural Information Processing Systems, 33*, 9855–9866.