Characterization of Temporal Trajectories of Autoimmune Atrophic Gastritis Using a Graph Representation Learning Pipeline

Chiara Sirtoli
Department of Management, Information and Production Engineering
University of Bergamo
Bergamo, Italy
chiara.sirtoli@unibg.it

Arianna Dagliati
Department of Electrical, Computer and Biomedical Engineering
University of Pavia
Pavia, Italy
arianna.dagliati@unipv.it

Marco Vincenzo Lenti
Department of Internal Medicine and Medical Therapeutics
University of Pavia
Pavia, Italy
marco.lenti@unipv.it

Antonio Di Sabatino
First Department of Internal Medicine
San Matteo Hospital Foundation
Pavia, Italy
a.disabatino@smatteo.pv.it

Antonio Ferramosca
Department of Management, Information and Production Engineering
University of Bergamo
Bergamo, Italy
antonio.ferramosca@unibg.it

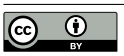
Daniele Pala
Department of Management, Information and Production Engineering
University of Bergamo
Bergamo, Italy
daniele.pala@unibg.it

Abstract

Autoimmune atrophic gastritis (AAG) is a chronic autoimmune disease affecting the stomach. Its diagnosis is based on serologic profiles, histological evaluations and esophagogastroduodenoscopies (EGDSs). AAG is characterized by a progressively increasing mucosal damage that can lead to several complications, however the specific temporal dynamics of this disease have not been fully characterized yet. This paper presents an exploratory study of the application of a graph representation learning method to model the AAG progression by identifying temporal phenotypes. We employed a procedure that combines Topological Data Analysis (TDA) and Minimum Spanning Tree filters, on a longitudinal dataset of patients that underwent multiple EGDSs, obtained from a multicentric Italian cohort. The dataset includes histological evaluations and clinical observations derived from text notes that were annotated in a previous study, with a total of 236 patients, 836 EGDSs performed, and 18 variables. Using our algorithm, we obtained a graph with weights representing the temporal progression. Five possible disease trajectories were identified through a Minimum Spanning Tree. Statistical tests and Cox Hazard Analysis demonstrated the existence of significant differences in the development of complications timing within these trajectories, proving that our approach is able to characterize the AAG evolution even in a relatively small dataset.

CCS Concepts

• **Applied computing** → *Health informatics.*



BCB '26, Rende (CS), Italy

ACM ISBN 979-8-4007-2653-8/26/06
https://doi.org/10.1145/3807503.3819432

Keywords

Autoimmune atrophic gastritis (AAG), Disease progression modelling, Topological Data Analysis (TDA)

ACM Reference Format:
Chiara Sirtoli, Arianna Dagliati, Marco Vincenzo Lenti, Antonio Di Sabatino, Antonio Ferramosca, and Daniele Pala. 2026. Characterization of Temporal Trajectories of Autoimmune Atrophic Gastritis Using a Graph Representation Learning Pipeline. In *17th ACM International Conference on Bioinformatics, Computational Biology and Health Informatics (BCB '26), June 30–July 03, 2026, Rende (CS), Italy*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3807503.3819432

1 Introduction

Characterized by the presence of anti-parietal cell (APCA) and anti-intrinsic factor (AIFA) antibodies, Autoimmune Atrophic Gastritis (AAG, also referred to as AIG) is a chronic, organ-specific condition characterized by an immune-mediated attack on the gastric oxyntic mucosa. This chronic, organ-specific autoimmune process results in deep structural changes to the stomach lining, leading to gastric atrophy and the loss of essential physiological functions, like reduction in gastric acid secretion and potential malabsorption issues.[1]

Due to the low prevalence in the general population (between 0,3% and 2,7%), AAG is considered a rare disease. AAG is more common in female (with a female-male ratio of 3:2) and older patients, particularly those above the age of 60. Moreover, the condition is diagnosed more frequently in patients with familiar predisposition and other autoimmune diseases.

Patients with AAG are often asymptomatic but can experience dietary or fasting-induced dyspepsia and reflux. Complications include reproductive health issues and iron-deficiency or pernicious anemia (from B12 malabsorption).

The diagnosis of AAG is done through serological tests, searching for example anti-parietal cell antibodies and anti-intrinsic factor

antibodies, esophagogastroduodenoscopies (EGDSs) and biopsies. No curative treatments are available for AAG. The progression of the disease is kept under observation via endoscopic surveillance and supplements are used to overcome deficiency of vitamin B12 and iron.

AAG is also categorized as a preneoplastic condition. It can lead to type I neuroendocrine tumors (NET) and gastric adenocarcinoma. In particular, patients with AAG result in an odd ratio of developing NET of about 11. Despite this, the prognosis is usually positive, with a low rate of metastasis in small NET. [1]

The association between AAG and NET is still unclear and there are some doubts about the possible risk factors. A study by Li et al. [5] proposed that Helicobacter pylori (H. pylori) infection, older age, higher vitamin B12 levels, gastric hyperplastic polyps could be preventive factors for NET in AAG patients [5]. Advanced age in particular could cause a potential diagnostic bias: the clinical significance of NET is likely to result in earlier identification of AAG, which might otherwise remain undetected. Conversely, the absence of NETs may cause a delayed diagnosis. Plus, studying the AAG-NET link is complicated by a lack of large, long-term patient datasets due to the disease's rarity.

To address these limitations, this study proposes the application of a Topological Data Analysis (TDA) framework, combined with a Minimum Spanning Tree (MST) filter, to analyze the progression of AAG. The primary objective is to determine whether patients follow distinct, identifiable trajectories characterized by different levels of risk and speed of progression. By mapping these clinical and temporal paths, we aim to clarify the role of age in NET occurrence, ultimately providing a more robust foundation for personalized endoscopic surveillance and the early detection of NET.

2 Methods

2.1 Data description and processing

Three datasets obtained from a multicentric Italian cohort were available for this study. In particular, the first dataset contains demographic and clinical data at the time of the diagnosis (like age at the diagnosis and drinking or smoking habits), while the second one reports longitudinal data about histological evaluations and text notes of patients' EGDS. The last dataset details the features present within the text notes from the second dataset. In a previous study, an encoder-decoder transformer model recovered the features contained within the notes [6]. We merged the second and the third datasets into a unique one composed by 379 patients, with a total of 759 EGDS performed, and 18 features. The features include 17 categorical variables (covering histological, endoscopic, and clinical indicators) and one temporal variable measuring the time between a specific EGDS and the patient's first one.

An analysis of the dataset's completeness revealed a significant proportion of missing data for the evaluated clinical and histological variables, with missingness rates ranging from 3.29% to 57.05%. To address the issue, the missing values were substituted by the value of the previous EGDS or set equal to zero in case of absence of previous values, following the indications provided by the clinicians, who stated that often pathological indicators are reported only if observed.

The combined dataset includes patients who have undergone from one up to four EGDS within a 256-month timeframe (around 20 years). The patients with only one EGDS were removed since they do not bring information about the temporal progression of the disease. At this point, the number of endoscopies reported is reduced to 632. The time distribution of the observations is illustrated in Figure S1(A) (presented in the supplementary materials) which reports the time interval of 20 years divided into 4 intervals of 5 years each. The observations made at the time of diagnosis are reported separately.

As reported in Figure S1(A), the distribution of the observations is imbalanced due to an higher concentration on EGDS performed in the first 64 months. To overcome this problem, we applied the Last Observation Carried Forward (LOCF) technique. More in particular, four total performed EGDS were imposed for each patient, i.e. in case of patients with a lower number of EGDS the last observation is replicated until the desired number is reached. The interval between the replicated EGDS and the previous one is kept equal to the one between the last two EGDS observed. Due to the high presence of observations in the first 128 months (around 10 years) in the

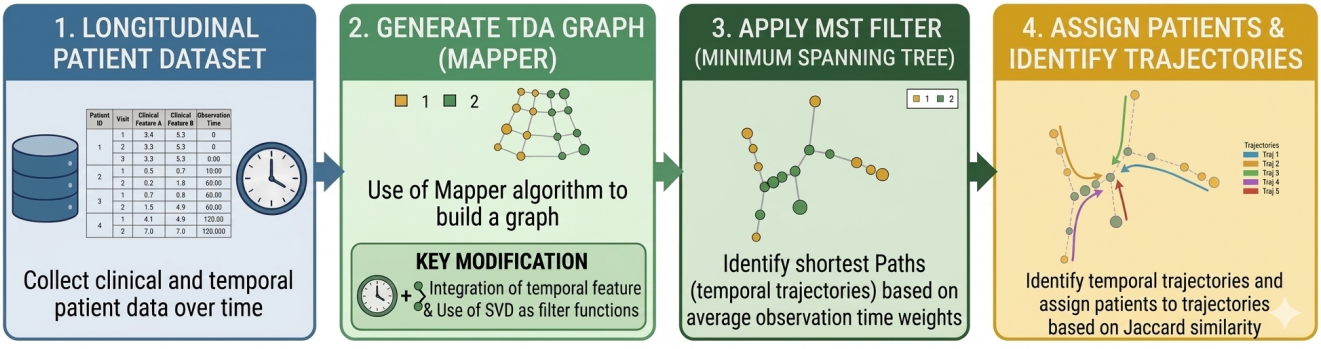


Figure 1: Schematic representation of the trajectory identification pipeline. The four-step framework encompasses longitudinal data collection, TDA graph construction via the Mapper algorithm (with integrated temporal features), MST filtering to extract the shortest temporal paths, and final patient assignment utilizing Jaccard similarity metrics. (Note: the figure was generated with the assistance of the Gemini AI model by Google).

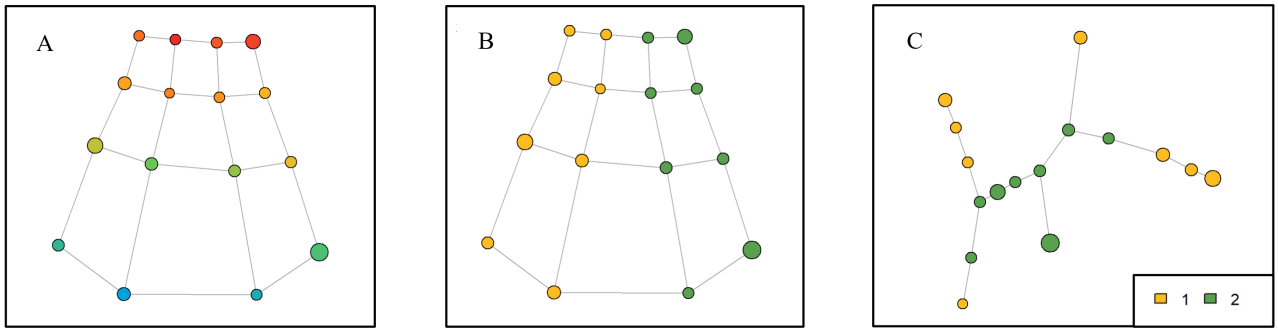


Figure 2: Topological Data Analysis (TDA) outputs and Minimum Spanning Tree (MST). (A) The Mapper graph with nodes colored according to the average observation time, illustrating the temporal progression of the disease. The color gradient ranges from light blue to red, indicating lower and higher average time values within the nodes, respectively. (B) The identical Mapper graph with nodes colored by their respective cluster assignment (Cluster 1 in yellow, Cluster 2 in green). (C) The resulting MST, colored by cluster.

combined dataset (see Figure S1(A)), only the EGDS reported in the first 120 months (10 years) were kept for the study.

At the end, the final dataset is composed of 236 patients and 836 EGDS (from 2 to 4 EGDS for patient) in a 10 year period. Figure S1(B) shows the distribution in time of the observations, in particular the time period is divided in four intervals of 30 months and the EGDSs at the diagnosis are counted separately. It is possible to note a distribution of observations more balanced than the pre-LOCF case. While the baseline period (i.e. time interval equal to zero) represents the largest number of observations, this is an expected outcome since it includes the initial diagnostic procedure that serves as the starting point for all patients.

2.2 Framework of TDA and MST

This study adopts an adapted version (illustrated in Figure 1) of a methodological framework introduced by Dagliati et al. [4] which combines a Topological Data Analysis (TDA) tool named Mapper, and a Minimum Spanning Tree (MST) filter. The original analysis code is publicly available at: https://github.com/trangdata/tdapseudotime. In particular, we applied the Mapper on our longitudinal dataset to generate a graph. We employed cosine similarity in conjunction with Singular Value Decomposition (SVD), extracting the first two singular vectors via the RSpectra package [https://github.com/yixuan/RSpectra] to serve as filter functions. Parameters were tuned to eliminate single-observation nodes. Moreover, a domain-driven heuristic focused on temporal coherence was followed with the aim to simultaneously maximize the chronological stratification of the clusters (ensuring a clear and granular temporal evolution across the identified clusters) and strictly enforce monotonic forward progression along the trajectories. This constraint prevented non-physiological backward loops, ensuring the extraction of biologically interpretable and temporally consistent clinical pathways. Differently from the original framework, we included the temporal feature to complement the cosine distance, which is unsuited to work purely with categorical data.

Following the identification of the topology, an MST filter was applied to discover the shortest paths, using the average observation time as edge weights. This process reveals trajectories that represent distinct disease progression pathways. Finally, as per Dagliati et al. [4], individual patients were assigned to these trajectories based on the Jaccard similarity, allowing for the identification of temporal phenotypes. The Jaccard similarity was computed between each patient's sequence (T_i) and each mined trajectory (T_d) using the intersection over union of their node sets:

$$J_{id}(T_i, T_d) = \frac{|T_i \cap T_d|}{|T_i \cup T_d|} \tag{1}$$

Each subject was then deterministically assigned to the trajectory with the highest similarity score.

3 Results

3.1 Topology and MST creation

Figure 2 shows the graph obtained from the Mapper algorithm and the resulting MST. The final topology was obtained by tuning the three core parameters of the Mapper: the number of clusters within each bin (defining the geometric scale), the degree of overlaps and the number of overlapping intervals (defining the resolution of the topology). We defined the optimal configuration using a geometric scale of 5, a 40% overlap and 4 interval bins.

The focus was to find the optimal parameter range capable of identifying clusters presenting distinct temporal profiles.

Differently from previous implementations of the framework [3, 4], the dimension of the dataset under analysis is significantly lower. This created a challenge to the TDA algorithm in the detection of multiple clusters that present a well-defined time distribution among them. While a visual comparison between the time-colored nodes (Figure 2(A)) and the cluster-colored nodes (Figure 2(B)) assignments suggests a coarse representation of temporal progression, the clusters capture distinct stages. As the statistical analysis reveals, the average and median time of the observations in each cluster are different. Specifically, Cluster 1 shows a mean

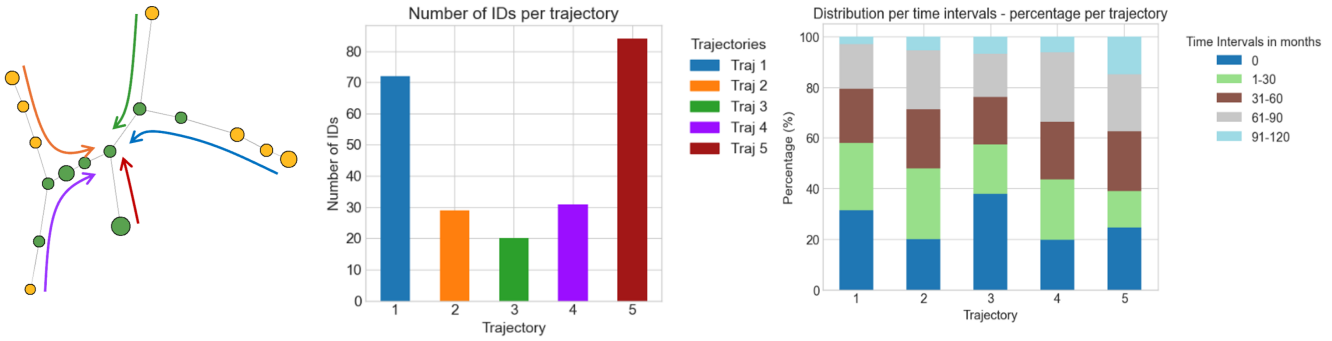


Figure 3: Visualization of the identified trajectories and their quantitative distribution. (Left) The Minimum Spanning Tree (MST) illustrating the five distinct temporal trajectories converging toward a shared, central node. (Middle) Bar chart detailing the number of patients assigned to each trajectory based on Jaccard similarity. (Right) Plot showing the balanced temporal distribution of clinical observations across the predefined time intervals for each trajectory.

time of 33.76 and a median time of 25.00, while Cluster 2 exhibits a higher mean time of 38.38 and a median time of 37.00. We further investigated these differences by applying a Kruskal-Wallis statistical test; the resulting p-value of 0.0477 confirms that the temporal distribution differences between the two clusters are statistically significant.

The temporal distinction between clusters is further maintained in the structure of the MST. In particular, the distribution of the nodes suggest a clear hierarchy: Cluster 1 occupies an internal position, while Cluster 2 is distributed across the terminal branches. This structural arrangement is also confirmed by the five distinct paths identified inside the MST. As shown in Figure 3 (Left), five temporal trajectories originating from the terminal nodes converge to a single, shared internal one.

3.2 Trajectories analysis

The MST (Figure 3 - Left) provides a topological visualization of patient states, illustrating how the five trajectories converge toward a central clinical hub. This pattern suggests that despite differing starting points, the disease progresses toward a common advanced state where patients become more clinically similar. To evaluate the framework's trajectory detection, we assessed the temporal distribution of EGDS observations and patient assignments via Jaccard similarity. A post-hoc validation confirmed unique assignments of each patient to the trajectories, satisfying the assumption of sample independence. The balanced observation distribution across time intervals (Figure 3 - Right) indicates that the trajectories reflect genuine, homogeneous clinical progression rather than data sparsity artifacts. Finally, patient distribution (Figure 3 - Middle) shows that trajectories 1 and 5 are the most populated, likely representing common clinical phenotypes, whereas the less populated paths may capture rarer progression pathways.

Table S1 (reported in the supplementary materials) summarizes the characteristics of all the observations (EGDS) in the dataset, stratified by the five temporal trajectories. For each variable, we report the distribution and frequency, while mean values and standard deviations are provided for the temporal variable (Time). Statistical significance of the differences across trajectories was evaluated using the Kruskal-Wallis test for continuous variables and the Chi-square test for categorical variables. While the mean observation time remains relatively homogeneous across all trajectories, most clinical and histological markers exhibit highly significant differences. Notably, trajectories 4 and 5 suggest high-risk evolution paths, accounting for the highest incidence of NETs (28.5% and 20.2%, respectively). On the other hand, trajectory 3 exhibited the lowest incidence of NET.

3.3 In-depth analysis of NET

Based on the clinical relevance of the AAG-NET relationship and the role of advanced age [2, 5], we further investigated NET progression across the trajectories. Using patient age at onset and at each EGDS, we extracted the minimum and mean age of first NET occurrence per trajectory. Specifically, the minimum and mean ages of occurrence were, respectively: 35 and 57.80 for Trajectory 1; 42 and 58.50 for Trajectory 2; 49 and 49.00 for Trajectory 3; 31 and 56.20 for Trajectory 4; and 46 and 68.45 for Trajectory 5. This revealed a strikingly premature diagnosis in trajectory 4 compared to the delayed clinical presentation observed in trajectory 5.

Table 1: Results of the Tukey's HSD pairwise comparisons detailing the mean differences in age of onset between trajectories. The p-values are adjusted for multiple comparisons to control the family-wise error rate.

Trajectories under analysis		Mean difference	p-value adjusted
1	2	0.7	0.9994
1	4	-1.6	0.9853
1	5	10.65	0.0661
2	4	-2.3	0.9813
2	5	9.95	0.3234
4	5	12.25	**0.0260**

The investigation on whether the age at NET occurrence varied across the temporal trajectories was done with a two-step statistical approach. First, a one-way ANOVA was performed to determine if at least one trajectory significantly differed from the others in terms of age at occurrence. The ANOVA reached statistical significance, yielding a p-value of 0.0174, thereby confirming that the age of onset is not uniform across clinical paths.

Subsequently, a Tukey's Honestly Significant Difference (HSD) post-hoc test was performed. This second step was crucial to search for specific differences, such as the significant temporal gap of twelve years in diagnosis between trajectories 4 and 5 (see Table 1).

Table 2: Risk of progression of each trajectory with follow-up time as time scale. Summarizes findings from the Cox proportional hazards model, including Hazard Ratio (HR), 95% Confidence Interval (CI), and corresponding p-value. Trajectory 3 is used as the reference baseline category for all comparisons.

Covariate (Trajectory)	Hazard Ratio (HR)	95% Confidence Interval (CI)	p-value
Trajectory 1	4.981	1.819 - 13.638	0.002
Trajectory 2	5.517	1.954 - 15.574	0.001
Trajectory 4	10.661	3.882 - 29.277	< 0.001
Trajectory 5	5.967	2.177 - 16.357	0.001

At this point, we conducted a Kaplan-Meier survival analysis and built a Cox Proportional Hazards model. Based on Tukey's HSD comparisons, Trajectories 4 and 5 represent advanced progression pathways. Consequently, the Kaplan-Meier survival analysis focuses on these three, using the stable, low-NET Trajectory 3 as a clinical baseline. In contrast, the Cox proportional hazards model includes all trajectories (1–5) to comprehensively evaluate hazard ratios across the spectrum, with Trajectory 3 as the reference. This model incorporates both patient age at each EGDS and follow-up time. The inclusion of follow-up time deliberately mitigates diagnostic age bias, helping determine if the advanced age in paths like Trajectory 5 reflects a true biological delay in tumorigenesis or merely the correlation between AAG diagnosis timing and NET detection.

Using follow-up time as the temporal variable for the analysis, the Kaplan-Meier curve reported in Figure 4 shows that the trajectory 3 maintains a consistently stable profile, indicating a minimal risk of NET. Conversely, trajectories 4 and 5 demonstrate a more rapid decline, especially trajectory 4. The higher risk assigned to trajectory 4 is further confirmed by the Cox regression analysis, which shows an elevated Hazard Ratio (HR) for this trajectory (see Table 2), alongside a lower but still significant risk for the trajectory 5.

These findings are fully confirmed also when the analysis is conducted using the patient's age as time scale. In fact, the risk stratification remains unchanged: trajectory 4 continues to display the highest HR, while trajectory 5 presents a lower risk that however remains statistically significant (see Table 3). The age-based Kaplan-Meier curves (Figure 5) also corroborate the considerations. Specifically, trajectory 4 shows a rapid decline at early age, while trajectory 5 declines more rapidly during advanced age.

Table 3: Risk of progression of each trajectory with patient's age as time scale. Summarizes findings from the Cox proportional hazards model, including Hazard Ratio (HR), 95% Confidence Interval (CI), and corresponding p-value. Trajectory 3 is used as the reference baseline category for all comparisons.

Covariate (Trajectory)	Hazard Ratio (HR)	95% Confidence Interval (CI)	p-value
Trajectory 1	6.544	2.379 - 18.003	< 0.001
Trajectory 2	6.202	2.195 - 17.525	0.001
Trajectory 4	11.494	4.177 - 31.624	< 0.001
Trajectory 5	7.691	2.801 - 21.118	< 0.001

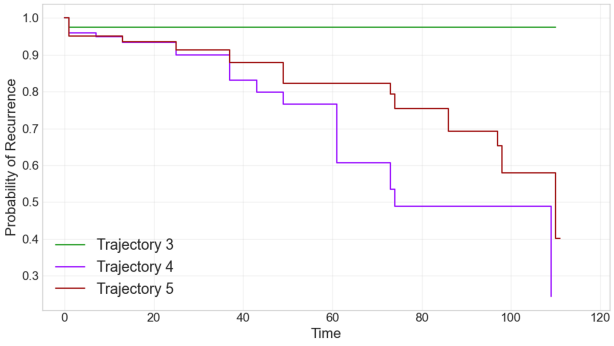


Figure 4: Kaplan-Meier survival analysis for NET occurrence. The curves illustrate the probability for patients assigned to Trajectories 3, 4, and 5 with follow-up time as time scale.

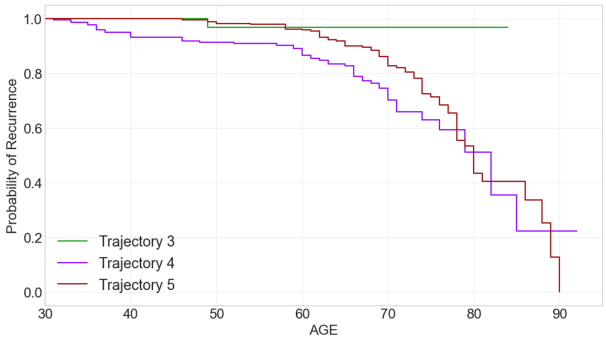


Figure 5: Kaplan-Meier survival analysis for NET occurrence. The curves illustrate the probability for patients assigned to Trajectories 3, 4, and 5 with patient's age as time scale.

4 Discussion

The application of a TDA and MST framework to AAG offers a novel perspective on its progression. A key methodological achievement is the framework's reliability on relatively small datasets; while TDA usually requires large cohorts, fine-tuning the geometric scale and topological resolution allowed the algorithm to effectively isolate meaningful clinical phenotypes. This represents a significant advancement for rare or newly identified diseases, which are often poorly characterized due to small sample sizes and limited longitudinal data. The stability of the five identified trajectories, supported by the balanced distribution of temporal observations and the structural consistency of the Minimum Spanning Tree (MST), underscores the robustness of this approach for rare diseases where large cohorts are difficult to recruit.

Despite the promising results, this study presents a few limitations that should be acknowledged. First, the imputation of missing data using the LOCF method inherently imposes an artificial stability across consecutive observations. Despite setting missing values to zero or carrying forward previous observations aligns with the clinical progression of the disease (where chronic alterations persist and undocumented features in EGDS reports indicate an absence of pathology), this approach inevitably imposes an artificial stability across the timeline. Consequently, the true temporal dynamics may be oversimplified, potentially masking minor clinical fluctuations. Second, although the proposed algorithmic framework demonstrates robust performance, the relatively small size of the dataset remains a constraint. This limited sample volume restricts the model's capacity to discover clusters that can truly represent different time periods.

Despite limitations, this study highlights significant clinical findings regarding NET progression across two temporal dimensions: follow-up time and patient age. While trajectory 3 remains a stable, low-risk phenotype with minimal events, trajectories 4 and 5 exhibit distinct high-risk profiles. Trajectory 4 represents an 'accelerated' pathway, characterized by exceptionally high Hazard Ratios in both follow-up (HR = 10.7) and age-based (HR = 11.2) analyses; its rapid increase in NET probability at earlier ages, even when normalized for follow-up time, demonstrates that this precocity is driven by intrinsic biological aggressiveness rather than diagnostic bias. Conversely, trajectory 5 exhibits a 'delayed' risk profile; although it carries a considerable risk during follow-up (HR = 5.7), the age-based Kaplan-Meier curve shows that this risk accumulates more slowly, precipitating predominantly at an advanced age. This delayed onset suggests a more protracted natural history, possibly influenced by hypothesized protective factors or representing a slower-moving variant of the disease.

Ultimately, shifting the analysis from chronological age to follow-up time successfully addressed potential diagnostic biases, confirming that the different paces of trajectories 4 and 5 reflect distinct biological behaviors of AAG rather than artifacts of NET detection timing.

5 Conclusion and Future Work

This study successfully demonstrated the validity of a TDA-based framework to partition the AAG clinical landscape, discovering five distinct trajectories with different progressions of neoplastic occurrence. We demonstrated that the risk of developing type I neuroendocrine tumors is not randomly distributed but follows specific paths: one characterized by rapid, younger-age progression (trajectory 4) and another by a delayed, older risk (trajectory 5).

Despite this, limitations remain mainly concerning the robustness of the experimental workflow. Strong dependence on LOCF and preprocessing choices may influence the time-based patterns rather than reflecting true disease behavior. Furthermore, the limited sample size and lack of independent validation render this work an exploratory, albeit promising, analysis requiring future validation with larger, more granular datasets.

Future work will integrate baseline lifestyle data (e.g., smoking and alcohol habits) to determine if external triggers drive the accelerated progression in trajectory 4 or support the stability in trajectory 3. Combining topological insights with these environmental variables will enable personalized monitoring protocols, optimizing endoscopic surveillance for AAG patients based on their specific evolutionary trajectory. Finally, future work will involve validation of the entire pipeline using an external cohort, which is not yet available; efforts are currently underway in collaboration with partner hospitals to collect additional patient data.

Acknowledgments

This work was funded by the National Plan for NRRP Complementary Investments (PNC) in the call for the funding of research initiatives for technologies and innovative trajectories in the health and care sectors (Directorial Decree n. 931 of 06-06-2022) – project n. PNC0000003 – AdvaNced Technologies for Human-centrEd Medicine (project acronym: ANTHEM). This work reflects only the authors' views and opinions, neither the Ministry for University and Research nor the European Commission can be considered responsible for them - CUP B53C22006700001.

References

[1] Chiara Castellana, Leonardo Henry Eusebi, Elton Dajti, Veronica Iascone, Amanda Vestito, Pietro Fusaroli, Lorenzo Fuccio, Antonietta D'Errico, and Rocco Maurizio Zagari. 2024. Autoimmune Atrophic Gastritis: A Clinical Review. 16, 7 (2024), 1310. doi:10.3390/cancers16071310

[2] Chuyan Chen, Yi Yang, Peng Li, and Haiyi Hu. 2023. Incidence of Gastric Neoplasms Arising from Autoimmune Metaplastic Atrophic Gastritis: A Systematic Review and Case Reports. 12, 3 (2023), 1062. doi:10.3390/jcm12031062

[3] Arianna Dagliati, Roberto Gatta, Alberto Malovini, Valentina Tibollo, Lucia Sacchi, Fidelia Cascini, Luca Chiovato, and Riccardo Bellazzi. 2022. A Process Mining Pipeline to Characterize COVID-19 Patients' Trajectories and Identify Relevant Temporal Phenotypes From EHR Data. 10 (2022), 815674. doi:10.3389/fpubh.2022.815674

[4] Arianna Dagliati, Nophar Geifman, Niels Peek, John H. Holmes, Lucia Sacchi, Riccardo Bellazzi, Seyed Erfan Sajjadi, and Allan Tucker. 2020. Using topological data analysis and pseudo time series to infer temporal phenotypes from electronic health records. 108 (2020), 101930. doi:10.1016/j.artmed.2020.101930

[5] Yan-Mei Li, Wen-Juan Guo, Chao Deng, Jie Luo, Yan-Fen Shi, Dan Zhu, Qi-Lu Wei, Ming-Gang Zhang, Shi-Yu Du, and Huang-Ying Tan. 2025. Risk assessment of type I gastric neuroendocrine tumors based on endoscopic and clinical features of autoimmune gastritis. 31, 41 (2025). doi:10.3748/wjg.v31.i41.111449

[6] Gabriele Sarti and Malvina Nissim. 2024. IT5: Text-to-text Pretraining for Italian Language Understanding and Generation. arXiv:2203.03759 [cs] doi:10.48550/arXiv.2203.03759