

ĐỒ ÁN TỐT NGHIỆP

**ỨNG DỤNG MÁY HỌC VECTOR HỖ TRỢ KÉP HẠT
CẦU CHO PHÂN LỚP ĐA NHÃN**

Ngành: **KHOA HỌC DỮ LIỆU**

Chuyên ngành: **KHOA HỌC DỮ LIỆU**

Giảng viên hướng dẫn : **TS. BÙI QUANG THỊNH**

Sinh viên thực hiện : Nguyễn Công Hùng

MSSV: 2186400229 Lớp: 21DKHA1

TP. Hồ Chí Minh, 2025

1. Lý do chọn đề tài

Trong kỷ nguyên bùng nổ dữ liệu hiện nay, các lĩnh vực nhận dạng mẫu, xử lý ảnh và y tế đang đối mặt với khối lượng thông tin ngày càng đa dạng và phức tạp. Điều này tạo ra nhu cầu cấp thiết về các phương pháp phân lớp đa nhãn tiên tiến và hiệu quả. Support Vector Machine (SVM) ¹ đã khẳng định vị thế là một trong những thuật toán học máy hiệu quả nhất, đặc biệt trong khai thác dữ liệu, xử lý ảnh và nhận dạng mẫu. Về nguyên lý, SVM hoạt động bằng cách xây dựng một siêu phẳng tối ưu nhằm phân tách dữ liệu thành hai lớp, đồng thời cực đại hóa margin (khoảng cách) đến các điểm dữ liệu gần nhất, còn gọi là support vectors. Nhờ nguyên lý này, SVM thường đạt được khả năng khái quát hóa cao, mang lại kết quả ấn tượng trong nhiều bài toán phân lớp và hồi quy khác nhau²⁻⁴.

Tuy nhiên, khi đối diện với dữ liệu có cấu trúc phức tạp, không đồng nhất hoặc phân bố không đều, SVM truyền thống bộc lộ nhiều hạn chế trong việc xác định siêu phẳng tối ưu. Đồng thời, chi phí tính toán tăng đáng kể khi kích thước dữ liệu mở rộng, làm giảm hiệu năng của SVM ở quy mô lớn⁵⁻⁷. Để khắc phục những hạn chế này, mô hình Twin Support Vector Machine (TSVM) ⁸ ra đời với ý tưởng sử dụng hai siêu phẳng phân tách thay vì một. Trong TSVM, mỗi siêu phẳng được xây dựng quanh một lớp dữ liệu, giúp giảm thiểu sai số do điểm ngoại lai và cải thiện hiệu năng tính toán. Về mặt lý thuyết, TSVM cho phép mô hình tập trung sát hơn vào từng lớp, nâng cao hiệu quả phân lớp trong nhiều tình huống thực tế ⁹⁻¹¹. Mặc dù vậy, TSVM vẫn tồn tại hạn chế khi xử lý dữ liệu phức tạp hoặc nhiễu, đồng thời đòi hỏi quá trình điều chỉnh tham số cẩn thận, bao gồm hệ số điều chỉnh sai số và vị trí siêu phẳng.

Nhằm tiếp tục nâng cao hiệu quả của TSVM, mô hình Granular Ball Twin Support Vector Machine (GBTSVM) ¹² được phát triển, kết hợp cơ chế phân lớp hai siêu phẳng của TSVM với ý tưởng phân cụm dữ liệu thành các granular ball (GB) ¹³. Phương pháp GB cho phép chia nhỏ không gian dữ liệu thành các nhóm đồng nhất, giúp mô hình tập trung vào từng vùng dữ liệu cục bộ và khai thác hiệu quả cấu trúc ẩn của dữ liệu. Kết quả là GBTSVM cải thiện đáng kể độ chính xác, đồng thời giảm chi phí tính toán nhờ việc phân rã dữ liệu thành các khối thông tin nhỏ hơn. Nhiều nghiên cứu đã mở rộng từ GBTSVM, tập trung vào tối ưu hóa tham số và kết hợp với các kỹ thuật tiền xử lý dữ liệu tiên tiến để nâng cao hiệu quả phân lớp với dữ liệu phức tạp ¹⁴⁻¹⁶. Tuy nhiên, việc chuyển từ bài toán phân lớp nhị phân sang đa nhãn—with yêu cầu gán nhiều nhãn cho mỗi mẫu—vẫn là thách thức lớn, đòi hỏi đổi mới cả về lý thuyết lẫn thực nghiệm.

Đề tài nghiên cứu của tôi hướng đến mục tiêu áp dụng và mở rộng GBTSVM cho bài toán phân lớp đa nhãn. Hiện tại đã có nghiên cứu phát triển việc áp dụng phân lớp đa nhãn cho GBTSVM thông qua việc thay đổi cấu trúc GB bằng sự kết hợp giữa Hierarchical Clustering và KMeans¹⁷. Tuy nhiên, hướng tiếp cận của tôi hoàn toàn khác biệt. Thay vì điều chỉnh cấu trúc GB, nghiên cứu này tập trung vào việc tích hợp khả năng phân lớp đa nhãn trực tiếp vào TSVM, trong đó GB chỉ đóng vai trò ổn định dữ liệu. Với cách tiếp cận này, mô hình mới không chỉ duy trì được ưu điểm của GBTSVM về độ chính xác và hiệu năng tính toán, mà còn mở rộng hiệu quả sang bài toán phân lớp đa nhãn, đáp ứng tốt hơn các yêu cầu của dữ liệu phức tạp và đa chiều trong thực tế.

2. Mục tiêu đề tài

Mục tiêu chung: Nghiên cứu này tập trung vào việc phát triển một phương pháp phân lớp đa nhãn dựa trên mô hình GBTSVM. Bài toán phân lớp đa nhãn - nơi mỗi đối tượng dữ liệu có thể đồng thời thuộc nhiều lớp khác nhau - đang trở thành một yêu cầu thiết yếu trong nhiều ứng dụng thực tế với dữ liệu phức tạp.

Mục tiêu cụ thể:

• Cải tiến mô hình GBTSVM cho phân lớp đa nhãn:

- Mở rộng kiến trúc GBTSVM, vốn được thiết kế cho phân lớp nhị phân, để xử lý hiệu quả bài toán đa nhãn - nơi mỗi mẫu dữ liệu có thể đồng thời thuộc nhiều lớp
- Xây dựng cơ chế tối ưu để cân bằng giữa độ chính xác phân lớp và chi phí tính toán, đặc biệt trong trường hợp dữ liệu có số lượng nhãn lớn.
- Thiết kế, mô hình hóa và triển khai thuật toán một cách chặt chẽ, có cơ sở khoa học rõ ràng.

• Đánh giá hiệu suất trên dữ liệu thực:

- Tiến hành thử nghiệm trên nhiều bộ dữ liệu thực tế đa nhãn với các đặc tính khác nhau số lượng nhãn, mức độ chồng chéo giữa các nhãn, kích thước dữ liệu.
- Áp dụng các tiêu chí đánh giá chuyên biệt cho bài toán đa nhãn như Hamming Loss, Subset Accuracy, Micro/Macro F1-score, để phân tích toàn diện hiệu năng của mô hình.
- So sánh độ chính xác, thời gian tính toán, khả năng khái quát hóa và tính ổn định giữa mô hình GBTSVM cải tiến và các phương pháp phân lớp truyền thống (SVM, TSVM, GBTSVM nguyên gốc, và các mô hình đa nhãn khác).

• Tối ưu hóa và đề xuất hướng phát triển mới:

- Nghiên cứu phương pháp tự động điều chỉnh tham số cho mô hình GBTSVM đa nhãn, bao gồm kích thước granular ball, hệ số phạt, và các tham số kernel.
- Phát triển các kỹ thuật tiền xử lý dữ liệu chuyên biệt để nâng cao hiệu quả của GBTSVM đa nhãn, như giảm chiều dữ liệu hoặc cân bằng phân phối nhãn.
- Xây dựng framework cho phép tích hợp mô hình vào các ứng dụng thực tế như phân loại văn bản, nhận dạng hình ảnh, và phân tích dữ liệu y học.

3. Tổng quan/Cơ sở lý luận

3.1 Giới thiệu về SVM và TSVM

3.1.1 Support Vector Machine (SVM)

SVM¹⁸ là một thuật toán học máy giám sát mạnh mẽ, dùng siêu phẳng để phân chia dữ liệu với khoảng cách biên lớn nhất. Các “vector hỗ trợ” – những điểm dữ liệu sát biên – quyết định vị trí siêu phẳng. Nhờ ánh xạ dữ liệu vào không gian chiều cao hơn thông qua các hàm kernel như Linear, Polynomial, RBF..., SVM có thể phân tách dữ liệu phức tạp một cách hiệu quả.

SVM xử lý tốt dữ liệu nhiều chiều, tổng quát hóa cao ngay cả khi tập dữ liệu nhỏ. Nhờ cơ chế kernel, nó linh hoạt trong việc phân tách các dạng dữ liệu khác nhau. Tuy nhiên, việc chọn kernel và tham số đòi hỏi kinh nghiệm, và SVM có thể tốn chi phí tính toán lớn khi dữ liệu quá lớn.

3.1.2 Twin Support Vector Machine (TSVM)

TSVM¹⁹ là một biến thể của SVM, trong đó hai siêu phẳng không song song được xây dựng, mỗi siêu phẳng gần một lớp dữ liệu hơn. Thay vì giải một bài toán tối ưu lớn như SVM truyền thống, TSVM tách thành hai bài toán tối ưu nhỏ hơn, nhờ đó có thể giảm bớt chi phí tính toán. Kết quả là mô hình linh hoạt trong việc phân lớp và có thể xử lý tốt nhiều dạng dữ liệu phức tạp²⁰.

Ưu điểm của TSVM nằm ở tốc độ huấn luyện nhanh hơn và khả năng tổng quát hóa tốt do chia nhỏ bài toán tối ưu. Mô hình này cũng ít nhạy với nhiễu và làm việc hiệu quả với dữ liệu có số chiều lớn. Tuy nhiên, việc lựa chọn tham số (như hàm kernel và hệ số phạt) đòi hỏi kinh nghiệm, và độ phức tạp vẫn tăng đáng kể khi kích thước dữ liệu quá lớn.

3.2 Giới thiệu về Granular Ball và GBTSVM

3.2.1 Granular Ball

Granular Ball (GB)²¹ là một phương pháp xử lý dữ liệu trong học máy và khai phá dữ liệu, tập trung nhóm các điểm dữ liệu tương đồng vào những “hạt cầu”. Mỗi quả cầu đại diện cho một vùng dữ liệu có chung đặc điểm, giúp giảm lược cấu trúc phức tạp và giảm nhiễu. Nhờ đó, các thuật toán học máy có thể xử lý dữ liệu nhanh hơn và chính xác hơn khi không phải làm việc trên toàn bộ dữ liệu thô^{22,23}.

Ưu điểm của GB là khả năng giảm kích thước dữ liệu, tối ưu hóa tính toán và tăng độ chính xác thông qua cơ chế gộp nhóm. Nó cũng giúp làm nổi bật các cụm dữ liệu quan trọng và hỗ trợ phát hiện nhiễu hoặc ngoại lệ hiệu quả. Tuy nhiên, việc lựa chọn tham số đòi hỏi thử nghiệm kỹ lưỡng, và hiệu suất có thể suy giảm trong các không gian dữ liệu cực lớn khi việc xác định ranh giới quả cầu trở nên khó khăn.

3.2.2 Granular Ball Twin Support Vector Machine (GBTSVM)

GBTSVM²⁴ kết hợp ý tưởng chia dữ liệu thành Granular Ball với cấu trúc hai siêu phẳng không song song của TSVM. Trước tiên, dữ liệu được gom thành những quả cầu tương đồng để giảm nhiễu và cô đọng thông tin. Sau đó, thuật toán Twin SVM được áp dụng lên tập dữ liệu đã được “nén” này, giúp mô hình tăng độ chính xác và tiết kiệm chi phí tính toán²⁵.

Lợi thế của GBTSVM nằm ở khả năng tổng quát hóa tốt, giữ được những ưu điểm của TSVM trong việc phân lớp linh hoạt, đồng thời nâng cao hiệu suất huấn luyện nhờ quá trình tiền xử lý dữ liệu bằng GB. Mặt khác, việc thiết lập các tham số như kích thước quả cầu, kernel, hệ số phạt trở nên phức tạp hơn và đòi hỏi tinh chỉnh kỹ lưỡng. Nếu dữ liệu quá lớn hoặc thiếu đồng nhất, khả năng gom nhóm bằng quả cầu có thể không đạt hiệu quả cao như mong muốn.

3.3 Tổng quan cơ chế kết hợp và cải tiến

Phương pháp tích hợp GBTSVM vào phân lớp đa nhãn bắt đầu với việc phân chia dữ liệu thành các hạt cầu (GB) có tính đồng nhất cao. Các GB này được hình thành dựa trên mức độ tương đồng giữa các điểm dữ liệu, giúp giảm thiểu nhiễu và làm nổi bật những vùng dữ liệu quan trọng. Từ nền tảng này, mô hình TSVM được áp dụng riêng biệt cho từng hạt cầu, tạo ra các siêu phẳng phân lớp tối ưu cho mỗi vùng dữ liệu cục bộ. Cách

tiếp cận này cho phép mô hình xử lý hiệu quả các không gian dữ liệu phức tạp, đặc biệt khi đối mặt với bài toán phân lớp đa nhãn.

Sự cải tiến chính của phương pháp nằm ở sự kết hợp hài hòa giữa hai thành phần chính: GB và TSVM. Trong đó, GB đóng vai trò phân tách dữ liệu thành các vùng cục bộ giàu thông tin, trong khi TSVM tập trung vào việc tối ưu hóa các siêu phẳng phân lớp cho từng vùng được xác định. Quá trình phân tách cục bộ này đặc biệt hiệu quả khi làm việc với dữ liệu có cấu trúc phức tạp hoặc yêu cầu độ chính xác cao trong môi trường đa nhãn. Các tham số quan trọng như bán kính hạt cầu, lựa chọn hàm kernel và hệ số điều chỉnh lỗi được TSVM tinh chỉnh phù hợp với đặc tính của từng vùng dữ liệu, từ đó nâng cao chất lượng dự đoán và tận dụng tối đa ưu điểm của cả hai thành phần.

4. Nội dung và phương pháp nghiên cứu

4.1. Đối tượng

Đối tượng nghiên cứu của đề tài bao gồm hai thành phần chính: bài toán phân lớp và thuật toán GBTSVM. Cụ thể, bài toán phân lớp được nghiên cứu từ góc độ lý thuyết lẫn ứng dụng, tập trung vào việc xử lý các trường hợp dữ liệu phức tạp, đặc biệt là trong bối cảnh phân lớp đa nhãn với nhiều cấu trúc ẩn. Mặt khác, thuật toán GBTSVM— vốn được thiết kế ban đầu cho bài toán phân lớp hai nhãn— sẽ được cải tiến và mở rộng để đáp ứng yêu cầu của các bài toán phân lớp đa nhãn, tận dụng cơ chế chia nhỏ không gian dữ liệu thành các GB đồng nhất nhằm xây dựng các siêu phẳng phân lớp cục bộ hiệu quả hơn.

4.2. Phạm vi nghiên cứu

Về dữ liệu: Nghiên cứu tập trung vào các bài toán phân lớp đa nhãn với dữ liệu có cấu trúc phức tạp, trong đó mỗi lớp chứa nhiều cụm con có đặc tính phân bố khác nhau. Các tập dữ liệu từ NDC, UCI, KEEL và các dữ liệu y tế công bố sẽ được sử dụng làm đối tượng thử nghiệm.

Về mô hình: Tập trung phát triển và đánh giá mô hình GBTSVM cho bài toán phân lớp đa nhãn, với trọng tâm đánh giá ba tiêu chí: độ chính xác phân lớp, thời gian tính toán và khả năng tổng quát hóa khi áp dụng trên các tập dữ liệu đã chọn.

4.3. Nội dung nghiên cứu

4.3.1 Khảo sát tài liệu và phân tích hiện trạng

Trong giai đoạn đầu tiên, Thu thập và tổng hợp các tài liệu, bài báo, công trình nghiên cứu liên quan đến bài toán phân lớp, bao gồm các mô hình SVM, TSVM và GBTSVM. Qua đó, đánh giá ưu, nhược điểm của từng phương pháp, đặc biệt về khả năng xử lý dữ liệu phức tạp và đa nhãn, nhằm tìm ra khoảng trống nghiên cứu cần cải tiến.

4.3.2 Phát triển mô hình GBTSVM cải tiến cho phân lớp đa nhãn

Dựa trên các kết quả khảo sát, nghiên cứu sẽ chuyển đổi mô hình GBTSVM từ bài toán hai nhãn sang phân lớp đa nhãn. Việc này bao gồm thiết kế kiến trúc mô hình mới, xây dựng hàm mục tiêu và thiết lập các ràng buộc tối ưu hóa phù hợp, đảm bảo khả năng phân biệt giữa các nhãn một cách hiệu quả.

4.3.3 Triển khai và thử nghiệm mô hình.

Lập trình và triển khai thuật toán trên môi trường thực tế sử dụng các công cụ phù hợp. Mô hình sẽ được thử nghiệm trên các bộ dữ liệu thực tế để điều chỉnh tham số và đánh giá hiệu suất qua các chỉ số như độ chính xác, thời gian tính toán và khả năng tổng quát hóa.

4.3.4 So sánh, đánh giá kết quả và đề xuất hướng phát triển

So sánh hiệu năng của mô hình cải tiến với các phương pháp truyền thống và các mô hình có chung hướng phát triển dựa trên các tiêu chí đã đề ra. Từ đó, rút ra những ưu điểm và hạn chế, đồng thời đề xuất các hướng nghiên cứu tiếp theo như cải tiến thuật toán tối ưu hóa hay mở rộng ứng dụng sang các lĩnh vực khác.

4.4. Phương pháp nghiên cứu

- Nghiên cứu lý thuyết: Tìm hiểu và tổng hợp các khái niệm, lý thuyết nền tảng về SVM, TSVM và GBTSVM từ tài liệu và báo cáo khoa học.
- Mô hình hóa toán học: Phát triển công thức, hàm mục tiêu và các ràng buộc cần thiết để xây dựng thuật toán GBTSVM.
- Mô phỏng và thực nghiệm: Triển khai thuật toán trên môi trường lập trình, sử dụng các bộ dữ liệu khác nhau để chạy mô phỏng và thu thập kết quả.

- Phân tích kết quả: Sử dụng các phương pháp thống kê và so sánh trực quan (đồ thị, biểu đồ) để đánh giá hiệu quả của thuật toán, từ đó rút ra các kết luận và đề xuất cải tiến.

5. Dự kiến kết quả

Về mặt lý thuyết

- **Nền tảng lý thuyết vững chắc:** Nghiên cứu tập trung vào cơ sở toán học của SVM, TSVM và cách tích hợp các khối “granular ball” để xây dựng GBTSVM, đảm bảo phân vùng dữ liệu và tìm siêu phẳng cực bộ tối ưu.
- **Mở rộng công thức toán học đa nhãn:** Phát triển hàm mục tiêu và ràng buộc mới cho bài toán đa nhãn, nâng cao khả năng xử lý dữ liệu phức tạp, đa chiều.
- **Giải pháp cho phân lớp phức tạp:** Kết hợp trực tiếp phân lớp đa nhãn vào cấu trúc GBTSVM, thay thế các phương pháp truyền thống One-vs-All hay One-vs-One, giúp mô hình linh hoạt hơn.
- **Tiền đề cho nghiên cứu nâng cao:** Tạo cơ sở cho các thuật toán tiên tiến như phân cụm động, học tăng cường, áp dụng với các kernel mới, phù hợp với yêu cầu dữ liệu lớn hiện nay.

Về mặt ứng dụng và phát triển

- **Chẩn đoán và phân loại bệnh y tế:** GBTSVM đa nhãn giúp xử lý tình huống bệnh nhân mắc nhiều bệnh đồng thời, hoặc một triệu chứng liên quan nhiều bệnh, nâng cao độ chính xác chẩn đoán.
- **Phát hiện nhãn phụ trong dữ liệu y tế:** Kết hợp cấu trúc hai siêu phẳng của TSVM và granular ball để nhận diện các bệnh đồng mắc hoặc triệu chứng ít gặp, hỗ trợ cải thiện quá trình chẩn đoán.
- **Định hướng nghiên cứu y học tương lai:** Mở rộng cho dự đoán tiến triển bệnh, phân tích nguy cơ, phản ứng thuốc; thúc đẩy hợp tác giữa chuyên gia y tế và khoa học dữ liệu trong bối cảnh y học chính xác.

Tài liệu tham khảo

1. Wang, Q. Support vector machine algorithm in machine learning. in *2022 IEEE international conference on artificial intelligence and computer applications (ICAICA)* 750–756 (IEEE, 2022).
2. Li, J., Li, Y., Song, J., Zhang, J. & Zhang, S. *Quantum Support Vector Machine for Classifying Noisy Data*. *IEEE Transactions on Computers* (IEEE, 2024).
3. Huang, M.-W., Chen, C.-W., Lin, W.-C., Ke, S.-W. & Tsai, C.-F. *SVM and SVM Ensembles in Breast Cancer Prediction*. *PloS one* vol. 12 e0161501 (Public Library of Science San Francisco, CA USA, 2017).

4. Chandra, M. A. & Bedi, S. S. Survey on SVM and their application in image classification. *Int. J. Inf. Technol.* **13**, 1–11 (2021).
5. Valero-Carreras, D., Alcaraz, J. & Landete, M. *Comparing Two SVM Models through Different Metrics Based on the Confusion Matrix*. *Computers & Operations Research* vol. 152 106131 (Elsevier, 2023).
6. Gatera, A., Kuradusenge, M., Bajpai, G., Mikeka, C. & Shrivastava, S. *Comparison of Random Forest and Support Vector Machine Regression Models for Forecasting Road Accidents*. *Scientific African* vol. 21 e01739 (Elsevier, 2023).
7. Bansal, M., Goyal, A. & Choudhary, A. *A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning*. *Decision Analytics Journal* vol. 3 100071 (Elsevier, 2022).
8. Tanveer, M., Rajani, T., Rastogi, R., Shao, Y. H. & Ganaie, M. A. Comprehensive review on twin support vector machines. *Ann. Oper. Res.* **339**, 1223–1268 (2024).
9. Moosaei, H., Ketabchi, S., Razzaghi, M. & Tanveer, M. Generalized Twin Support Vector Machines. *Neural Process. Lett.* **53**, 1545–1564 (2021).
10. Tanveer, M., Tiwari, A., Choudhary, R. & Ganaie, M. A. Large-scale pinball twin support vector machines. *Mach. Learn.* **111**, 3525–3548 (2022).
11. Tanveer, M., Ganaie, M. A., Bhattacharjee, A. & Lin, C.-T. Intuitionistic fuzzy weighted least squares twin SVMs. *IEEE Trans. Cybern.* **53**, 4400–4409 (2022).
12. Quadir, A., Sajid, M. & Tanveer, M. Granular Ball Twin Support Vector Machine. *IEEE Trans. Neural Netw. Learn. Syst.* 1–10 (2024) doi:10.1109/TNNLS.2024.3476391.
13. Xia, S., Dai, X., Wang, G., Gao, X. & Giem, E. An efficient and adaptive granular-ball generation method in classification problem. *IEEE Trans. Neural Netw. Learn. Syst.* **35**, 5319–5331 (2022).
14. Quadir, A. & Tanveer, M. Granular ball twin support vector machine with pinball loss function. *IEEE Trans. Comput. Soc. Syst.* (2024).
15. Ganaie, M. A. & Ahire, V. Granular Ball Twin Support Vector Machine with Universum Data. Preprint at <https://doi.org/10.48550/arXiv.2412.03375> (2024).
16. Quadir, A., Sajid, M., Tanveer, M. & Suganthan, P. N. Enhanced Feature Based Granular Ball Twin Support Vector Machine. Preprint at <https://doi.org/10.48550/arXiv.2410.05786> (2025).
17. Ganaie, M. A., Ahire, V. & Girard, A. Granular Ball K-Class Twin Support Vector Classifier. Preprint at <https://doi.org/10.48550/arXiv.2412.05438> (2024).
18. Guido, R., Ferrisi, S., Lofaro, D. & Conforti, D. An Overview on the Advancements of Support Vector Machine Models in Healthcare Applications: A Review. *Information* **15**, (2024).
19. Huang, H., Wei, X. & Zhou, Y. *An Overview on Twin Support Vector Regression*. *Neurocomputing* vol. 490 80–92 (Elsevier, 2022).
20. Thapa, S., Adhikari, S., Ghimire, A. & Aditya, A. Feature selection based twin-support vector machine for the diagnosis of Parkinson's disease. in *2020 IEEE 8th R10 humanitarian technology conference (R10-HTC)* 1–6 (IEEE, 2020).
21. Xia, S. *et al.* Granular-ball fuzzy set and its implement in svm. *IEEE Trans. Knowl. Data Eng.* (2024).
22. Xia, S., Zheng, S., Wang, G., Gao, X. & Wang, B. Granular ball sampling for noisy label classification or imbalanced classification. *IEEE Trans. Neural Netw. Learn. Syst.* **34**, 2144–2155 (2021).
23. Xie, J., Kong, W., Xia, S., Wang, G. & Gao, X. An efficient spectral clustering algorithm based on granular-ball. *IEEE Trans. Knowl. Data Eng.* **35**, 9743–9753 (2023).
24. Zhao, L., Zhang, Z., Liu, W. & Lang, G. GBTWSVM: Granular-Ball Twin Support Vector Machine. in *Rough Sets* (eds. Hu, M. *et al.*) vol. 14839 238–251 (Springer Nature Switzerland, Cham, 2024).
25. Xia, S. *et al.* *Gbsvm: An Efficient and Robust Support Vector Machine Framework via Granular-Ball Computing*. *IEEE Transactions on Neural Networks and Learning Systems* (IEEE, 2024).

Tiến độ thực hiện

| Tuần | 1-2 | 3-4 | 5-6 | 7-8 | 9-10 | 11-12 |
|--|-----|-----|-----|-----|------|-------|
| Dự kiến nội dung thực hiện | | | | | | |
| Nghiên cứu tài liệu, tìm hiểu về SVM, TSVM và Granularball | | | | | | |
| Tổng hợp kiến thức và khái quát cho GBTSVM | | | | | | |
| Xác định mục tiêu, phạm vi và xây dựng cơ sở lý thuyết | | | | | | |
| Mô hình hóa toán học và thiết kế kiến trúc mô hình | | | | | | |
| Bắt đầu xây dựng và triển khai thuật toán | | | | | | |
| Tiếp tục triển khai thuật toán và tinh chỉnh | | | | | | |
| Thử nghiệm mô hình trên dữ liệu thực tế | | | | | | |
| So sánh, đánh giá kết quả và phân tích dữ liệu | | | | | | |
| Bổ sung, hoàn thiện báo cáo và chuẩn bị bảo vệ đồ án | | | | | | |

Tp.HCM, ngày 17 tháng 03 năm 2025
SINH VIÊN THỰC HIỆN
(Ký và ghi rõ họ tên)

Hùng

Nguyễn Công Hùng

XÁC NHẬN CỦA BỘ MÔN

GIẢNG VIÊN HƯỚNG DẪN
(Ký và ghi rõ họ tên)

TS. Bùi Quang Thịnh