



ĐỀ CƯƠNG 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: 2186400267 Lớp: 21DKHA1

MỤC LỤC

Mục lục	1
Chương 1. Lý do chọn đề tài	2
Chương 2. Mục tiêu đề tài	4
Chương 3. Tổng quan/cơ sở lý luận	6
3.1 Giới thiệu về SVM và TSVM	6
3.1.1 Support Vector Machine (SVM)	6
3.1.2 Twin Support Vector Machine (TSVM)	6
3.2 Giới thiệu về Granular Ball và GBTSVM	6
3.2.1 Granular Ball	6
3.2.2 Granular Ball Twin Support Vector Machine (GBTSVM)	7
3.3 Tổng quan cơ chế kết hợp và cải tiến	7
Chương 4. Nội dung và phương pháp nghiên cứu	8
4.1 Đối tượng nghiên cứu	8
4.2 Phạm vi nghiên cứu	8
4.3 Dự kiến kết quả	8
4.3.1 Khảo sát tài liệu và phân tích hiện trạng	8
4.3.2 Phát triển mô hình GBTSVM cải tiến cho phân lớp đa nhãn. . .	8
4.3.3 Triển khai và thử nghiệm mô hình	8
4.3.4 So sánh, đánh giá kết quả và đề xuất hướng phát triển	9
4.4 Phương pháp nghiên cứu	9
Chương 5. Dự kiến kết quả	10
TÀI LIỆU THAM KHẢO	12

Chương 1. Lý do chọn đề tài

Trong những năm gần đây, khối lượng dữ liệu đa dạng và phức tạp trong các lĩnh vực như nhận dạng mẫu, xử lý ảnh và y tế không ngừng gia tăng, kéo theo nhu cầu cấp thiết về các phương pháp phân lớp đa nhãn tiên tiến. Trong bối cảnh này, Support Vector Machine (SVM) [1] đã được công nhận là một trong những thuật toán học máy phổ biến và hiệu quả nhất, đặc biệt trong các ứng dụng khai thác dữ liệu, xử lý ảnh và nhận dạng mẫu. Về mặt lý thuyết, SVM xác định một siêu phẳng tối ưu (optimal hyperplane) nhằm phân tách dữ liệu thuộc hai lớp khác nhau, đồng thời đảm bảo khoảng cách lớn nhất đến các điểm dữ liệu biên (support vectors). Nhờ vậy, SVM thường đạt được độ khái quát hóa cao, cho kết quả ổn định trong nhiều bài toán phân lớp và hồi quy [2]–[4]. Tuy nhiên, khi đối mặt với dữ liệu có cấu trúc phức tạp, không đồng nhất hoặc phân bố không đều, mô hình SVM truyền thống bắt đầu bộc lộ hạn chế trong việc tìm kiếm siêu phẳng tối ưu. Bên cạnh đó, khi kích thước dữ liệu mở rộng, chi phí tính toán gia tăng đáng kể, khiến SVM mất lợi thế về tốc độ và hiệu năng ở quy mô lớn [5]–[7].

Để khắc phục những hạn chế của SVM, mô hình Twin Support Vector Machine (TSVM) [8] đã được đề xuất với ý tưởng sử dụng hai siêu phẳng phân tách thay vì chỉ một. Trong TSVM, mỗi siêu phẳng được xây dựng xung quanh một lớp dữ liệu, giúp giảm thiểu sai số do các điểm ngoại lai đồng thời cải thiện hiệu năng tính toán so với SVM thông thường. Về mặt lý thuyết, TSVM tận dụng cấu trúc hai siêu phẳng để cho phép mô hình tập trung “gần” hơn vào từng lớp, từ đó nâng cao hiệu quả phân lớp trong nhiều kịch bản dữ liệu thực nghiệm [9]–[11]. Dù vậy, TSVM vẫn có những hạn chế, đặc biệt khi dữ liệu có độ phức tạp cao hoặc chứa nhiều nhiễu. Mô hình này thường đòi hỏi việc điều chỉnh tham số cẩn thận, bao gồm cả hệ số điều chỉnh sai số và vị trí siêu phẳng; nếu không được tối ưu hóa phù hợp, TSVM có thể không đạt được kết quả mong muốn khi áp dụng cho các bài toán phức tạp.

Nhằm tiếp tục cải thiện hiệu quả của TSVM, mô hình Granular Ball Twin Support Vector Machine (GBTSVM) [12] ra đời, 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 này 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 tận dụng triệt để cấu trúc ẩn của dữ liệu. Kết quả là GBTSVM có thể cải thiện độ chính xác, đồng thời giảm chi phí tính toán do dữ liệu được phân rã thành các khối thông tin nhỏ hơn. Bên cạnh đó, một số nghiên cứu đã mở rộng từ mô hình GBTSVM nhằm giải

quyết các bài toán phức tạp hơn, tập trung vào việc tối ưu hóa tham số và kết hợp với các kỹ thuật tiền xử lý dữ liệu tiên tiến để cải thiện hiệu quả phân lớp trong trường hợp dữ liệu có cấu trúc phức tạp [14]–[16]. Dù những nỗ lực này đã cho kết quả khả quan đối với bài toán phân lớp hai nhãn, việc chuyển sang bài toán phân lớp đa nhãn – với yêu cầu gán nhiều nhãn cho mỗi mẫu dữ liệu – vẫn là một thách thức lớn, đòi hỏi cải tiến sâu về cả mặt lý thuyết và thực nghiệm.

Chính vì vậy, đề tài nghiên cứu của tôi được hình thành với mục tiêu áp dụng và mở rộng ứng dụng của GBTSVM vào bài toán phân lớp đa nhãn. Hơn thế nữa, trong thời điểm hiện tại cũng đã có bài báo nghiên cứu phát triển việc áp dụng phân lớp đa nhãn cho GBTSVM bằng cách thông qua việc thay đổi cấu trúc GB bằng sự kết hợp giữa Hierarchical Clustering và KMeans [17], tuy nhiên hướng nghiên cứu của tôi áp dụng lại hoàn toàn khác biệt. Thay vì điều chỉnh cấu trúc GB, nghiên cứu này sẽ tập trung vào việc tích hợp khả năng phân lớp đa nhãn trực tiếp vào TSVM, trong đó granular ball chỉ đảm nhận vai trò ổn định dữ liệu. Nhờ vậy, mô hình mới không chỉ giữ được ưu điểm của GBTSVM về độ chính xác và hiệu năng tính toán, mà còn mở rộng ứng dụng sang bài toán phân lớp đa nhãn, đáp ứng hiệu quả các yêu cầu của dữ liệu phức tạp và đa chiều.

Chương 2. Mục tiêu đề tài

Mục tiêu chung: Phát triển phương pháp phân lớp đa nhãn dựa trên mô hình Granular Ball Twin Support Vector Machine (GBT SVM), góp phần giải quyết hiệu quả các bài toán có dữ liệu đa lớp và cấu trúc phức tạp. Việc mở rộng mô hình này nhằm đảm bảo khả năng khai thác tối đa thông tin ẩn trong dữ liệu, giảm thiểu sai số, cũng như cải thiện tốc độ và khả năng khái quát hóa của mô hình.

Mục tiêu cụ thể:

- **Cải tiến mô hình GBT SVM cho phân lớp đa nhãn:**

- Điều chỉnh kiến trúc và thuật toán của GBT SVM, vốn được thiết kế cho bài toán phân lớp hai nhãn, để có thể gán nhiều nhãn cho mỗi đối tượng dữ liệu.
- Đảm bảo khả năng xử lý tốt dữ liệu có cấu trúc phức tạp, không đồng nhất và có độ chiều cao.
- 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:**

- Thực hiện các thí nghiệm trên nhiều bộ dữ liệu thực tế, đặc biệt là những bộ dữ liệu có tính đa nhãn rõ rệt và cấu trúc phức tạp, để kiểm chứng tính hiệu quả của mô hình.
- So sánh độ chính xác, thời gian tính toán, khả năng khái quát hóa và tính ổn định giữa mô hình GBT SVM cải tiến và các phương pháp phân lớp truyền thống (SVM, TSVM, GBT SVM nguyên gốc, và các mô hình đa nhãn khác).
- Sử dụng các tiêu chí đánh giá chuyên sâu như độ chính xác, độ tin cậy, F1-score, v.v. để phân tích một cách toàn diện kết quả mô hình.

- **Tối ưu hóa và đề xuất hướng phát triển mới:**

- Tìm kiếm cách lựa chọn và hiệu chỉnh tham số (chẳng hạn kích thước “hạt”, hệ số phạt, v.v.) giúp mô hình đạt hiệu suất cao nhất trên các bộ dữ liệu thử nghiệm.
- Đề xuất giải pháp mở rộng hoặc tích hợp các kỹ thuật tiền xử lý, như kỹ thuật giảm chiều dữ liệu hoặc lọc nhiễu, nhằm nâng cao hơn nữa chất lượng phân lớp.

- Phác thảo định hướng ứng dụng mô hình cải tiến trong những lĩnh vực đòi hỏi phân lớp đa nhãn, như nhận dạng hình ảnh.

Chương 3. Tổng quan/cơ sở lý luận

3.1. Giới thiệu về SVM và TSVM

3.1.1. Support Vector Machine (SVM)

SVM [18] 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)

TSVM [19] 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 tạp [20].

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 (GBTWSVM)

GBTWSVM [24] 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 toán [25].

Lợi thế của GBTWSVM 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

Cơ chế tích hợp GBTWSVM vào phân lớp đa nhãn khởi đầu bằng việc phân chia dữ liệu thành các GB có tính chất đồng nhất cao. Các hạt cầu này được hình thành dựa trên mức độ tương đồng của dữ liệu, qua đó giúp loại bỏ nhiễu và làm nổi bật các vùng dữ liệu quan trọng. Trên cơ sở đó, TSVM sẽ được áp dụng cục bộ cho từng hạt cầu để xây dựng các siêu phẳng phân lớp tối ưu cho mỗi vùng riêng biệt. Bằng cách này, mô hình có thể xử lý hiệu quả các không gian dữ liệu phức tạp, đặc biệt trong các bài toán đa nhãn.

Điểm cải tiến nằm ở việc kết hợp hài hòa giữa hai thành phần: GB đảm nhiệm việc tách lớp dữ liệu, định hình các vùng cục bộ giàu thông tin; TSVM tập trung vào tối ưu hóa siêu phẳng phân lớp cho từng vùng này. Sự phân tách cục bộ này đặc biệt hữu ích khi dữ liệu có cấu trúc phức tạp hoặc đòi hỏi độ chính xác cao trong bối cảnh nhiều nhãn. Các tham số như bán kính hạt cầu, lựa chọn kernel và hệ số phạt khi đưa vào sẽ được TSVM tinh chỉnh tương ứng với từng vùng, giúp nâng cao chất lượng dự đoán và tận dụng tốt ưu thế của từng thành phần.

Chương 4. Nội dung và phương pháp nghiên cứu

4.1. Đối tượng nghiên cứu

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: Tập trung vào các bài toán phân lớp đa nhãn, trong đó mỗi lớp chứa nhiều cụm con với đặc tính phân bố khác nhau, đặc biệt trên các tập dữ liệu NDC, UCI, KEEL.

Về mô hình: Nghiên cứu phát triển và đánh giá mô hình GBTSVM, áp dụng cho các tập dữ liệu đã nêu để kiểm tra hiệu quả về độ chính xác, thời gian tính toán và khả năng tổng quát hóa.

4.3. Dự kiến kết quả

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 (như Python, MATLAB...). 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.

Chương 5. Dự kiến kết quả

Về mặt lý thuyết

- **Xây dựng nền tảng lý thuyết vững chắc:** Nghiên cứu cơ sở toán học và lý thuyết để tải tiến được GBTSVM một cách hợp lý, làm rõ cơ chế phân chia dữ liệu cũng như cách thức xây dựng các siêu phẳng cục bộ.
- **Mở rộng công thức toán học cho bài toán phân lớp đa nhãn:** Phát triển các biểu thức tối ưu mới, chú trọng vào việc gán nhiều nhãn cho một mẫu dữ liệu, qua đó nâng cao tính linh hoạt và khả năng xử lý của mô hình khi đối diện với cấu trúc dữ liệu phức tạp.
- **Cung cấp giải pháp mới cho bài toán phân lớp phức tạp:** Phát triển hướng tiếp cận hiện đại cho các bài toán đòi hỏi gán nhiều nhãn, đóng góp thêm một lựa chọn về mặt lý thuyết trong lĩnh vực nhận dạng mẫu, xử lý ảnh, phân tích dữ liệu y tế, ...
- **Tiền đề cho các nghiên cứu nâng cao:** Tạo cơ sở khoa học cho việc phát triển các thuật toán tiên tiến khác ví dụ, phân cụm động hoặc học tăng cường, đáp ứng tốt hơn nhu cầu trong thời đại dữ liệu lớn và đa dạng.

Về mặt thực nghiệm

- **Triển khai thành công mô hình GBTSVM cải tiến:** Hiện thực hóa mô hình GBTSVM cho phân lớp đa nhãn, đảm bảo tính ổn định và hiệu năng cao trên cả dữ liệu tổng hợp lẫn dữ liệu thực tế.
- **Đạt kết quả vượt trội so với các phương pháp hiện có:** Thông qua các bộ tiêu chí đánh giá như F1-score, độ chính xác, độ nhạy, ..., mô hình mới được kỳ vọng thể hiện ưu thế về độ chính xác, thời gian huấn luyện và khả năng tổng quát hóa so với SVM truyền thống hay TSVM cũ.
- **Phân tích chuyên sâu hiệu năng:** Tiến hành thí nghiệm định lượng và định tính để xác định mức độ phù hợp của mô hình với nhiều loại dữ liệu, đặc biệt là dữ liệu có chiều lớn, phân bố không đều hoặc chứa nhiễu, qua đó khẳng định tính linh hoạt và tiềm năng ứng dụng rộng rãi.

Về mặt ứng dụng và phát triển

- **Mở rộng phạm vi ứng dụng:** Kết quả nghiên cứu có thể được áp dụng trong nhiều lĩnh vực đòi hỏi năng lực phân lớp đa nhãn, chẳng hạn nhận dạng

đối tượng trong ảnh, phân tích y tế đa chiều, phân tích văn bản và hệ thống khuyến nghị.

- **Nền tảng cho các thuật toán học máy tiên tiến:** Nghiên cứu này không chỉ có ý nghĩa về mặt lý thuyết mà còn tạo tiền đề cho việc phát triển các phương pháp cải tiến khác, ví dụ tích hợp mô hình Granular Ball với các kỹ thuật học sâu (deep learning) hoặc áp dụng cho phân cụm động.
- **Định hướng cho nghiên cứu tương lai:** Mô hình GBTSVM cải tiến có thể được mở rộng sang nhiều dạng bài toán khác như học bán giám sát, học trực tuyến, học tăng cường và thúc đẩy hợp tác đa ngành, giải quyết những thách thức dữ liệu phức tạp trong kỷ nguyên số.

TÀI LIỆU THAM KHẢO

- [1] Q. Wang, “Support vector machine algorithm in machine learning,” in *2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA)*, 2022, pp. 750–756. DOI: [10.1109/ICAICA54878.2022.9844516](https://doi.org/10.1109/ICAICA54878.2022.9844516).
- [2] J. Li, Y. Li, J. Song, J. Zhang, and S. Zhang, “Quantum support vector machine for classifying noisy data,” *IEEE Transactions on Computers*, vol. 73, no. 9, pp. 2233–2247, 2024. DOI: [10.1109/TC.2024.3416619](https://doi.org/10.1109/TC.2024.3416619).
- [3] M.-W. Huang, C.-W. Chen, W.-C. Lin, S.-W. Ke, and C.-F. Tsai, “Svm and svm ensembles in breast cancer prediction,” *PLOS ONE*, vol. 12, no. 1, pp. 1–14, Jan. 2017. DOI: [10.1371/journal.pone.0161501](https://doi.org/10.1371/journal.pone.0161501). [Online]. Available: <https://doi.org/10.1371/journal.pone.0161501>.
- [4] M. A. Chandra and S. S. Bedi, “Survey on svm and their application in imageclassification,” *International Journal of Information Technology*, vol. 13, no. 5, pp. 1–11, 2021, ISSN: 2511-2112. DOI: [10.1007/s41870-017-0080-1](https://doi.org/10.1007/s41870-017-0080-1). [Online]. Available: <https://doi.org/10.1007/s41870-017-0080-1>.
- [5] D. Valero-Carreras, J. Alcaraz, and M. Landete, “Comparing two svm models through different metrics based on the confusion matrix,” *Computers Operations Research*, vol. 152, p. 106 131, 2023, ISSN: 0305-0548. DOI: <https://doi.org/10.1016/j.cor.2022.106131>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0305054822003616>.
- [6] A. Gatera, M. Kuradusenge, G. Bajpai, C. Mikeka, and S. Shrivastava, “Comparison of random forest and support vector machine regression models for forecasting road accidents,” *Scientific African*, vol. 21, e01739, 2023, ISSN: 2468-2276. DOI: <https://doi.org/10.1016/j.sciaf.2023.e01739>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2468227623001953>.
- [7] M. Bansal, A. Goyal, and A. Choudhary, “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, p. 100 071, 2022, ISSN: 2772-6622. DOI: <https://doi.org/10.1016/j.dajour.2022.100071>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S2772662222000261>.

- [8] M. Tanveer, T. Rajani, R. Rastogi, Y. H. Shao, and M. A. Ganaie, “Comprehensive review on twin support vector machines,” *Annals of Operations Research*, vol. 339, no. 3, pp. 1223–1268, 2024, ISSN: 1572-9338. DOI: [10.1007/s10479-022-04575-w](https://doi.org/10.1007/s10479-022-04575-w). [Online]. Available: <https://doi.org/10.1007/s10479-022-04575-w>.
- [9] H. Moosaei, S. Ketabchi, M. Razzaghi, and M. Tanveer, “Generalized twin support vector machines,” *Neural Processing Letters*, vol. 53, no. 2, pp. 1545–1564, 2021, ISSN: 1573-773X. DOI: [10.1007/s11063-021-10464-3](https://doi.org/10.1007/s11063-021-10464-3). [Online]. Available: <https://doi.org/10.1007/s11063-021-10464-3>.
- [10] M. Tanveer, A. Tiwari, R. Choudhary, and M. A. Ganaie, “Large-scale pinball twin support vector machines,” *Machine Learning*, vol. 111, no. 10, pp. 3525–3548, 2022, ISSN: 1573-0565. DOI: [10.1007/s10994-021-06061-z](https://doi.org/10.1007/s10994-021-06061-z). [Online]. Available: <https://doi.org/10.1007/s10994-021-06061-z>.
- [11] M. Tanveer, M. A. Ganaie, A. Bhattacharjee, and C. T. Lin, “Intuitionistic fuzzy weighted least squares twin svms,” *IEEE Transactions on Cybernetics*, vol. 53, no. 7, pp. 4400–4409, 2023. DOI: [10.1109/TCYB.2022.3165879](https://doi.org/10.1109/TCYB.2022.3165879).
- [12] A. Quadir, M. Sajid, and M. Tanveer, “Granular ball twin support vector machine,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–10, 2024. DOI: [10.1109/TNNLS.2024.3476391](https://doi.org/10.1109/TNNLS.2024.3476391).
- [13] S. Xia, X. Dai, G. Wang, X. Gao, and E. Giem, “An efficient and adaptive granular-ball generation method in classification problem,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 5319–5331, 2024. DOI: [10.1109/TNNLS.2022.3203381](https://doi.org/10.1109/TNNLS.2022.3203381).
- [14] A. Quadir and M. Tanveer, “Granular ball twin support vector machine with pinball loss function,” *IEEE Transactions on Computational Social Systems*, pp. 1–10, 2024. DOI: [10.1109/TCSS.2024.3411395](https://doi.org/10.1109/TCSS.2024.3411395).
- [15] M. A. Ganaie and V. Ahire, *Granular ball twin support vector machine with universum data*, 2024. arXiv: [2412.03375](https://arxiv.org/abs/2412.03375) [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2412.03375>.
- [16] A. Quadir, M. Sajid, M. Tanveer, and P. N. Suganthan, *Enhanced feature based granular ball twin support vector machine*, 2025. arXiv: [2410.05786](https://arxiv.org/abs/2410.05786) [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2410.05786>.
- [17] M. A. Ganaie, V. Ahire, and A. Girard, *Granular ball k-class twin support vector classifier*, 2024. arXiv: [2412.05438](https://arxiv.org/abs/2412.05438) [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2412.05438>.

- [18] R. Guido, S. Ferrisi, D. Lofaro, and D. Conforti, “An overview on the advancements of support vector machine models in healthcare applications: A review,” *Information*, vol. 15, no. 4, p. 235, 2024.
- [19] H. Huang, X. Wei, and Y. Zhou, “An overview on twin support vector regression,” *Neurocomputing*, vol. 490, pp. 80–92, 2022, ISSN: 0925-2312. DOI: <https://doi.org/10.1016/j.neucom.2021.10.125>. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0925231222003101>.
- [20] S. Thapa, S. Adhikari, A. Ghimire, and A. Aditya, “Feature selection based twin-support vector machine for the diagnosis of parkinson’s disease,” in *2020 IEEE 8th R10 Humanitarian Technology Conference (R10-HTC)*, 2020, pp. 1–6. DOI: [10.1109/R10-HTC49770.2020.9356984](https://doi.org/10.1109/R10-HTC49770.2020.9356984).
- [21] S. Xia, X. Lian, G. Wang, X. Gao, Q. Hu, and Y. Shao, “Granular-ball fuzzy set and its implement in svm,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, no. 11, pp. 6293–6304, 2024. DOI: [10.1109/TKDE.2024.3419184](https://doi.org/10.1109/TKDE.2024.3419184).
- [22] S. Xia, S. Zheng, G. Wang, X. Gao, and B. Wang, “Granular ball sampling for noisy label classification or imbalanced classification,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 4, pp. 2144–2155, 2023. DOI: [10.1109/TNNLS.2021.3105984](https://doi.org/10.1109/TNNLS.2021.3105984).
- [23] J. Xie, W. Kong, S. Xia, G. Wang, and X. Gao, “An efficient spectral clustering algorithm based on granular-ball,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 9, pp. 9743–9753, 2023. DOI: [10.1109/TKDE.2023.3249475](https://doi.org/10.1109/TKDE.2023.3249475).
- [24] L. Zhao, Z. Zhang, W. Liu, and G. Lang, “Gbtwsvm: Granular-ball twin support vector machine,” in *Rough Sets*, M. Hu, C. Cornelis, Y. Zhang, P. Lingras, D. Ślezak, and J. Yao, Eds., Cham: Springer Nature Switzerland, 2024, pp. 238–251, ISBN: 978-3-031-65665-1.
- [25] S. Xia, X. Lian, G. Wang, X. Gao, J. Chen, and X. Peng, “Gbsvm: An efficient and robust support vector machine framework via granular-ball computing,” *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15, 2024. DOI: [10.1109/TNNLS.2024.3417433](https://doi.org/10.1109/TNNLS.2024.3417433).

Tiến độ thực hiện

Dự kiến nội dung thực hiện	Tuần							
	1-2	3-4	5-6	7-8	8-9	10-11	12	13
Nghiên cứu tài liệu, tìm hiểu về SVM, TSVM và Granular ball								
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								

Bảng 1: Dự kiến tiến độ công việc

Tp.HCM, ngày 26 tháng 02 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)



Bùi Quang Thịnh