

# Giải quyết bài toán phân loại với dữ liệu mất cân bằng dựa trên mạng GAN

Sinh viên                      Đinh Tuấn Anh & Trần Nam Bá & Ngô Đức Bảo

Giảng viên dạy              PGS. TS Huỳnh Trung Hiếu & Lưu Giang Nam

Hội đồng phản biện      PGS. TS Huỳnh Trung Hiếu & TS. Nguyễn Chí Kiên & TS. Bùi Thanh Hùng & Lưu Giang Nam

## Tóm tắt nội dung

Mất cân bằng dữ liệu là một trong những hiện tượng phổ biến của bài toán phân loại, nếu hiện tượng mất cân bằng nghiêm trọng xảy ra thường dẫn tới dự báo kém chính xác trên nhóm thiểu số bởi đa phần kết quả dự báo ra thường thiên về nhóm đa số và rất kém trên nhóm thiểu số. Trong đề tài này, chúng em sẽ tiến hành thử nghiệm xử lý vấn đề mất cân bằng dữ liệu bằng phương pháp sử dụng mạng GAN để tạo mẫu cho nhóm thiểu số trong tập dữ liệu. Huấn luyện mô hình với bộ dữ liệu ban đầu và bộ dữ liệu mới với các mẫu được sinh ra từ mạng GAN. So sánh kết quả đạt được để thấy được hiệu quả của phương pháp này.

## Bố cục bài báo cáo

Cấu trúc bài báo cáo gồm 5 phần:

1. Giới thiệu
2. Các kiến thức liên quan
3. Phương pháp
4. Kết quả thử nghiệm
5. Kết luận và cải tiến trong tương lai.

## Tài liệu

- [1] Kim, J., Jeong, K., Choi, H., Seo, K. (2020). GAN-Based Anomaly Detection In Imbalance Problems. In: Bartoli, A., Fusiello, A. (eds) *Computer Vision – ECCV 2020 Workshops. ECCV 2020. Lecture Notes in Computer Science*, 12540. Springer, Cham. 10.1007/978-3-030-65414-611