

## Một phương pháp ứng dụng mẫu nén và học máy để phát hiện Flycam trong môi trường có chồng lấn với tín hiệu WiFi

Trần Vũ Kiên, Đỗ Anh Tú, Nguyễn Hải Quân, Nguyễn Lê Cường\*

Trường Đại học Điện lực.

\*Email: cuongnl@epu.edu.vn

Nhận bài: 12/8/2022; Hoàn thiện: 04/10/2022; Chấp nhận đăng: 10/10/2022; Xuất bản: 28/10/2022.

DOI: <https://doi.org/10.54939/1859-1043.j.mst.82.2022.70-80>

### TÓM TẮT

Bài báo này nghiên cứu đề xuất một giải pháp phát hiện tín hiệu WiFi lẫn trong tín hiệu vô tuyến phát ra từ Flycam. Giải pháp được đề xuất dựa trên phương pháp học máy và kỹ thuật lấy mẫu nén trong trường hợp phổ tần của tín hiệu thu được từ Flycam và từ các thiết bị sử dụng sóng WiFi bị chồng lấn lên nhau tại điểm thu tín hiệu. Để thực hiện điều đó, một ma trận lấy mẫu nén đặc biệt được thiết kế với mục đích chỉ thu lấy đoạn tín hiệu có chứa tín hiệu đồng bộ để sau đó trích chọn các đặc trưng và đưa vào hệ thống học máy nhằm phát hiện và loại bỏ được đoạn tín hiệu WiFi có lẫn trong tín hiệu thu được. Một hệ thống máy thu cũng được xây dựng trên nền tảng vô tuyến cấu hình mềm (SDR) để thử nghiệm phương pháp này. Kết quả tính toán và thử nghiệm cho thấy, giải pháp được đề xuất giúp cải thiện tốc độ trong việc loại bỏ tín hiệu WiFi có lẫn trong Flycam.

**Từ khóa:** Lấy mẫu nén; Nhiễu WiFi; Học máy.

### 1. ĐẶT VẤN ĐỀ

Trong những năm gần đây, việc sử dụng máy bay không người lái, cụ thể là Flycam một cách dễ dàng, có tính tự phát, không làm thủ tục xin cấp phép bay với cơ quan có thẩm quyền, hoặc bay vào khu vực cấm, đã gây ra các vấn đề ảnh hưởng đến an toàn hàng không, an toàn dân sinh và an ninh quốc gia. Do đó, việc phát hiện và cảnh báo sớm sự xuất hiện của các Flycam để có các phản ứng kịp thời là rất cần thiết. Để giải quyết vấn đề này, đã có nhiều nghiên cứu theo các cách tiếp cận khác nhau như: phân tích âm thanh, phân tích hình ảnh thu được từ camera kết hợp với các thuật toán học máy và trí tuệ nhân tạo để cải thiện độ chính xác và hiệu quả của việc phát hiện, sử dụng radar chủ động và phân tích dựa trên thu tín hiệu vô tuyến trao đổi giữa Flycam và trạm điều khiển mặt đất. Mỗi cách tiếp cận có những ưu nhược điểm và phạm vi ứng dụng khác nhau.

Hầu hết các Flycam sử dụng ba loại tín hiệu vô tuyến: tín hiệu điều khiển từ xa, tín hiệu đo từ xa, và tín hiệu video. Tín hiệu đo từ xa chủ yếu truyền về vị trí, độ cao và trạng thái hoạt động của UAV có điều chế giống với tín hiệu điều khiển. Tín hiệu video là tín hiệu truyền hình ảnh thu được từ camera gắn trên Flycam về trạm mặt đất và thường được phát liên tục trong quá trình bay của Flycam, tuy nhiên, các tín hiệu video thường sử dụng kỹ thuật điều chế OFDM và hoạt động trong dải tần 2,4 GHz hoặc 5,8 GHz Flycam trùng dải tần với WiFi, gây khó khăn cho việc xử lý nhận dạng khi tại vị trí thu tín hiệu vô tuyến từ Flycam có các thiết bị WiFi hoạt động.

Một số nghiên cứu được đề xuất để loại bỏ tín hiệu WiFi trong xử lý và nhận dạng Flycam. Các kỹ thuật dự đoán độ dài tín hiệu OFDM kết hợp với các đặc trưng của phổ công suất làm đầu vào cho mô hình SVM và KNN được áp dụng trong [1] để phân loại tín hiệu. Ezuma và cộng sự trong [2] dựa vào độ rộng băng thông để loại bỏ tín hiệu WiFi. Bài báo [3] phân biệt tín hiệu WiFi và Flycam bằng cách sử dụng 15 đặc tính thống kê trong miền thời gian và tần số, kết hợp với tính toán sóng mang con hiệu dụng, tuy nhiên, các phương pháp trên cần tính toán toàn bộ dữ liệu vô tuyến thu được dẫn đến khối lượng tính toán lớn.

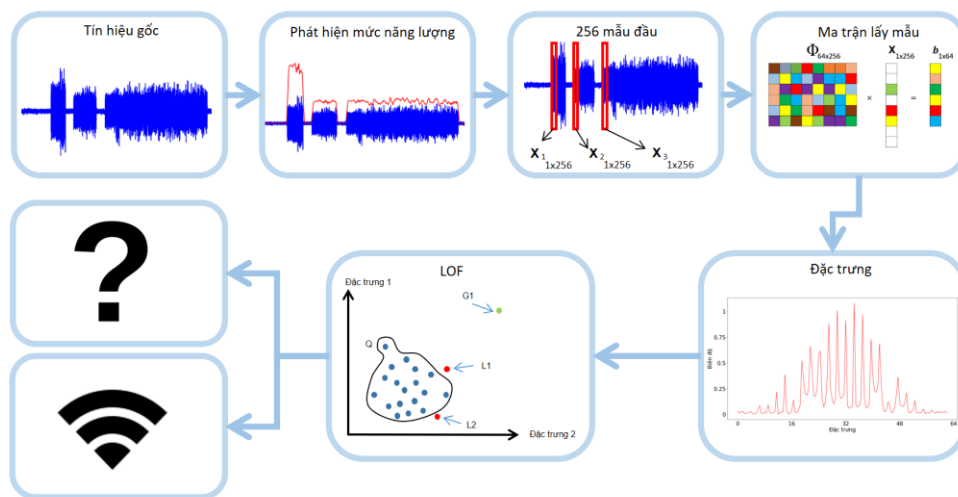
Các nghiên cứu trên chưa tính đến trường hợp tín hiệu Flycam và tín hiệu WiFi chia sẻ cùng một dải tần và các tính toán nhận biết đều phải tính đến toàn bộ dữ liệu thu được dẫn đến khó

khăn đáp ứng yêu cầu thời gian thực. Để giải quyết vấn đề này, bài báo đề xuất một phương pháp nhận biết tín hiệu WiFi trong tín hiệu gốc dựa trên lấy mẫu nén và học máy. Cụ thể, đối với tín hiệu thu được trong miền thời gian, đầu tiên, sự xuất hiện của tín hiệu được phát hiện bằng cách sử dụng kỹ thuật phát hiện mức năng lượng [4]. Tiếp theo, trích xuất tín hiệu đồng bộ của tín hiệu WiFi nếu có từ những đoạn tín hiệu trước đó. Sau đó, thực hiện lấy mẫu nén đoạn tín hiệu trích xuất với một ma trận lấy mẫu được thiết kế dựa trên các đặc trưng của tín hiệu WiFi. Cuối cùng, tín hiệu mẫu nén được sử dụng làm dữ liệu đầu vào cho mô hình học máy để nhận dạng tín hiệu WiFi và sử dụng làm điều kiện để loại bỏ tín hiệu WiFi trong tín hiệu gốc.

Phần còn lại của bài báo được trình bày như sau: Phần 2 trình bày giải pháp được đề xuất. Phần 3 trình bày quá trình mô phỏng và thử nghiệm. Phần 4 thảo luận đánh giá hiệu quả của phương pháp được đề xuất với các kịch bản. Cuối cùng, phần 5 trình bày các kết luận.

## 2. ĐỀ XUẤT GIẢI PHÁP

Trong thực tế, có những trường hợp Flycam và thiết bị thu phát WiFi hoạt động trong cùng một băng tần, tuy nhiên, tùy theo khoảng cách giữa Flycam, thiết bị WiFi và địa điểm thu chặn tín hiệu mà tín hiệu thu được bị ảnh hưởng theo chiều hướng khác nhau. Trong trường hợp Flycam bay trong vùng hoạt động của các nguồn phát tín hiệu WiFi (ví dụ: Flycam bay qua tòa nhà), router WiFi hoạt động theo giao thức CSMA/CA để tránh xung đột khi có thiết bị khác hoạt động cùng băng tần. Vì vậy, mặc dù hoạt động cùng băng tần nhưng tín hiệu thu được từ Flycam và WiFi không chồng lấn trong miền thời gian. Trường hợp Flycam không bay trong vùng hoạt động của nguồn phát tín hiệu WiFi nhưng cả 2 loại thiết bị nằm trong vùng phủ của bộ thu chặn tín hiệu, điều này dẫn đến việc thu tín hiệu có sự chồng lấn ở miền phổ và miền thời gian, trường hợp này đặc biệt gây khó khăn cho việc xử lý tín hiệu. Trong cả 2 trường hợp việc loại bỏ tín hiệu WiFi đều hỗ trợ quá trình xử lý tín hiệu thu chặn được từ Flycam dễ dàng hơn.



**Hình 1.** Mô hình hệ thống xử lý tín hiệu thu được.

Bài báo này đề xuất mô hình xử lý loại bỏ tín hiệu WiFi khỏi tín hiệu thu được ban đầu như trong hình 1. Tín hiệu gốc sau khi được thu từ bộ thu sẽ tiếp tục được phát hiện mức năng lượng để tách ra làm các đoạn tín hiệu riêng biệt. Sau đó, các đoạn tín hiệu này tiếp tục được lấy mẫu nén sử dụng một ma trận lấy mẫu nén nhằm mục đích làm nổi bật ra các đặc trưng của tín hiệu WiFi, và được sử dụng làm dữ liệu đầu vào cho mô hình học máy phân loại có phải WiFi hay không.

### 2.1. Phát hiện mức năng lượng

Tín hiệu vô tuyến thu được được biểu diễn  $x(t) = \sum_{i=1}^n \alpha_i \psi_i(t)$ , trong đó,  $\Psi$  là một ma trận để biểu diễn tín hiệu có các cột là  $i$ . Các cột của ma trận  $\Psi$  là một chuỗi số được tạo từ kết quả

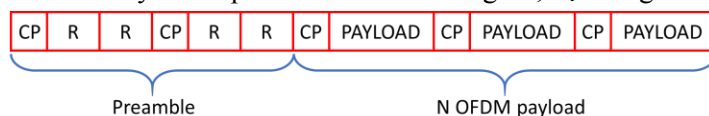
hàm tính năng lượng của tín hiệu vô tuyến thu được. Sự xuất hiện của tín hiệu vô tuyến trong không gian có thể được nhận biết thông qua hàm tính năng lượng như trong biểu thức sau:

$$E_S = \langle x(n), x(n) \rangle = \sum_{i=0}^N |x(n)|^2 \quad (1)$$

Kết quả của hàm tính năng lượng cho thấy, biên độ năng lượng tại những vị trí có tín hiệu sẽ cao hơn so với vị trí không xuất hiện tín hiệu, kết quả này được sử dụng để loại bỏ những đoạn nhiễu nền nhằm cải thiện thời gian tính toán cho hệ thống. Trước khi thực hiện lấy mẫu tính toán, ngưỡng nhiễu nền được thiết lập và có giá trị  $\eta$ , một khoảng  $\sigma$  được lựa chọn để phân biệt tín hiệu so với nhiễu nền.

## 2.2. Lấy mẫu nén

WiFi sử dụng đa truy nhập phân chia theo tần số trực giao (OFDM), cấu trúc một khung truyền OFDM bao gồm 2 phần là preamble và các OFDM symbols được biểu diễn trong hình 2. Phần đầu tín hiệu preamble trong tín hiệu WiFi thường kéo dài 2 symbols dùng để đồng bộ tín hiệu bên phía thu, các OFDM symbols phía sau sẽ chứa thông tin, nội dung của bản tin.



Hình 2. Cấu trúc khung truyền OFDM.

Đa số các kỹ thuật đồng bộ sử dụng preamble cho khung truyền OFDM được thực hiện bằng kỹ thuật do Schmidl và Cox đề xuất được trình bày trong [5]. Preamble được tạo ra bằng 1 symbol có 2 nửa tín hiệu giống nhau trong miền thời gian bằng cách truyền chuỗi PN trong miền tần số ở các kênh sóng mang con chẵn, các kênh sóng mang con lẻ sẽ không truyền bất kỳ tín hiệu nào. Hầu hết các bộ thu phát WiFi hoạt động có băng thông 20 MHz hoặc 40 MHz. Đối với các kênh WiFi có băng thông là 20 MHz sẽ chứa 64 kênh sóng mang con, 1 symbol có tối đa 80 mẫu trong miền thời gian, trong đó, 16 mẫu đầu là phần CP và 64 mẫu là dữ liệu. Do tín hiệu preamble của WiFi chỉ truyền với băng thông 20 MHz và kéo dài trong 2 symbol, điều này có nghĩa là, số lượng mẫu trong miền thời gian có tối đa 160 mẫu. Trong bài báo sử dụng 256 mẫu đầu tiên là đủ cho 2 symbol của preamble WiFi, đảm bảo thời gian tính toán cho mỗi tín hiệu một cách nhanh chóng nhưng vẫn mang đủ đặc trưng.

Có thể nhận thấy rằng, đoạn tín hiệu chứa preamble WiFi chiếm số lượng rất ít nếu tính trên toàn bộ dữ liệu WiFi khi giao tiếp và có thể coi là các tín hiệu thừa trên toàn bộ miền dữ liệu của nó. Hơn nữa, chúng lại là đặc trưng của tín hiệu WiFi, đối với một đoạn tín hiệu vô tuyến chỉ cần nhận diện được đoạn preamble thì có thể kết luận đoạn tín hiệu đó là tín hiệu WiFi và có thể được loại bỏ trong quá trình tính toán nhận dạng Flycam.

Từ các phân tích trên, 256 mẫu đầu tiên của mỗi đoạn tín hiệu trong miền thời gian sẽ được lấy, đối với mỗi đoạn tín hiệu gồm 256 mẫu đầu này sẽ xây dựng được các cột tương ứng của ma trận  $\Psi$ .

Khi đó, bài toán lấy mẫu nén đối với tín hiệu preamble được xây dựng như sau.

$$b = \Phi x = \Phi \Psi \alpha \quad (2)$$

Trong đó,  $b$  là tín hiệu sau lấy mẫu nén,  $\Psi$  là một ma trận biểu diễn tín hiệu thừa,  $\Phi$  là ma trận lấy mẫu nén. Trong bài báo này, ma trận lấy mẫu nén  $\Phi$  được tạo thành từ ma trận Fourier. Thông thường ma trận này được tạo thành từ ma trận ngẫu nhiên sẽ cho kết quả khôi phục lại tín hiệu với độ chính xác cao, nhưng các ma trận như vậy yêu cầu không gian lưu trữ lớn cũng như độ phức tạp tính toán cao trong quá trình khôi phục lại tín hiệu thừa được lấy mẫu bởi các ma trận này [6]. Đối với bài toán cụ thể được đề xuất trong bài báo, tín hiệu OFDM từ các thiết bị WiFi có một số thông số đặc biệt biết trước như băng thông hoạt động, số kênh con. Dựa vào các thông số này bài báo đề xuất một ma trận lấy mẫu đơn giản và có hiệu quả cao khi được áp dụng

trong trường hợp đặc biệt này. Ma trận lấy mẫu nén đề xuất được tạo thành từ các hàng của ma trận Fourier  $N \times N$  [7].

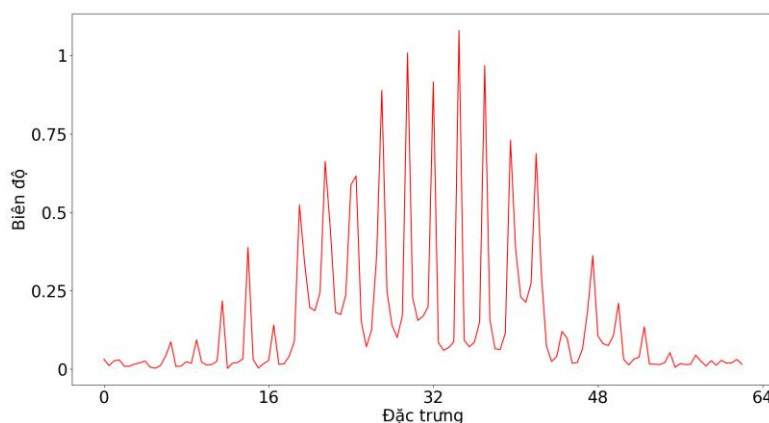
Giả thiết  $F_0, F_1, \dots, F_{N-1}$  là các hàng của ma trận Fourier  $N \times N, F^{(N)}$

$$F^{(N)} = (F_0 F_1 \dots F_{N-1}). \quad (3)$$

Ma trận lấy mẫu  $\Phi$  được chọn là ma trận riêng phần của ma trận  $F^{(N)}$  với  $M$  hàng. Khi đó  $M = \{m_0, m_1, \dots, m_r\} \subset \{0, 1, \dots, N-1\}$  với các thành phần bị loại bỏ tương ứng với các vị trí tần số ở đây có sóng mang con lè trong OFDM. Có thể xác định ma trận Fourier riêng phần tương ứng với  $M$  là

$$F_M^{(N)} = (F_{m_0} F_{m_1} \dots F_{m_r}). \quad (4)$$

Khi đó, kết quả của biểu thức 2 chính là tín hiệu lấy mẫu nén 256 mẫu đầu tiên của tín hiệu WiFi và được thể hiện như trong hình 3. Tín hiệu này mang đặc trưng cho phần preamble của tín hiệu WiFi và chỉ gồm 64 mẫu, tạo thuận lợi khi được đưa vào một mô hình học máy như được trình bày trong mục 2.3.

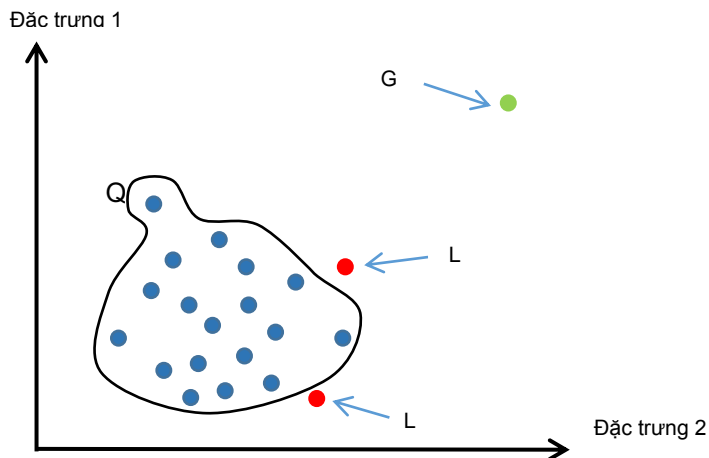


**Hình 3.** Đặc trưng từ mẫu thưa thu được.

### 2.3. Mô hình học máy

Các nghiên cứu trong [8-11] đã chỉ ra việc sử dụng tín hiệu vô tuyến thu được để phân loại Flycam với WiFi hoặc giữa các loại Flycam khác nhau dựa trên các mô hình học máy là khả thi. Nhằm mục đích tự động phân loại tín hiệu WiFi và phần còn lại, bài báo sẽ lựa chọn một mô hình học máy phù hợp với dữ liệu thu được và với việc trích chọn 64 mẫu đặc trưng trong 256 mẫu đầu của tín hiệu như trình bày trong các phần trên sẽ giúp làm giảm khối lượng tính toán, đơn giản hóa mô hình và giảm thời gian huấn luyện của hệ thống.

Trong học máy, phân loại một lớp truyền thống có thể chia thành 2 nhóm chính là: phương pháp dựa trên khoảng cách và phương pháp dựa trên mật độ [12]. Trong số các phương pháp đó một phương pháp nổi tiếng có thể giải quyết hiệu quả vấn đề về dữ liệu nhiều chiều đó là Local Outlier Factor (LOF) [13] đã được Ezuma và cộng sự trình bày trong [8]. Việc giải quyết dữ liệu càng nhiều chiều thì khối lượng tính toán càng tăng vậy nên việc chọn các phương pháp quá phức tạp cho bài toán là chưa khả thi. Do đó, giải pháp đề xuất trong bài báo lựa chọn ứng dụng mô hình học máy LOF là một trong những bộ học máy truyền thống nhằm tối giản hóa độ phức tạp và tăng tốc độ tính toán. Những mẫu chứa đặc trưng đã trích chọn ở phần trên được đưa vào mô hình LOF huấn luyện nhằm phân loại được tín hiệu đầu vào. Thuật toán LOF được Ezuma và cộng sự trình bày trong [8] gán cho dữ liệu đầu vào một trọng số để quyết định việc dữ liệu này có phải là dữ liệu ngoại lai (outlier) hay không bằng cách tính khoảng cách từ điểm dữ liệu đó tới cụm các điểm lân cận, một dữ liệu được gọi là dữ liệu ngoại lai khi nó nằm tách biệt khỏi cụm.



**Hình 4.** Biểu diễn của một tập dữ liệu 2 chiều.

Hình 4 biểu diễn một tập dữ liệu 2 chiều bao gồm dữ liệu gốc và hai loại dữ liệu ngoại lai là dữ liệu ngoại lai cục bộ (local outlier) và dữ liệu ngoại lai toàn cục (global outlier). Trong đó tập dữ liệu ban đầu có 1 lớp là Q và các điểm dữ liệu mới đến là (L1, L2, G1) và khi đó L1 và L2 sẽ là “dữ liệu ngoại lai cục bộ” của Q và G1 sẽ là “dữ liệu ngoại lai toàn cục” của Q.

Trong giải pháp được đề xuất ở đây, những dữ liệu được cho là WiFi sẽ được huấn luyện thành các điểm dữ liệu gốc sau đó những dữ liệu còn lại sẽ được phân ra là các dữ liệu ngoại lai khác không được cho là WiFi, cụ thể trong bài báo là các mẫu tín hiệu Flycam.

### 3. THỬ NGHIỆM

#### 3.1. Các trường hợp thử nghiệm

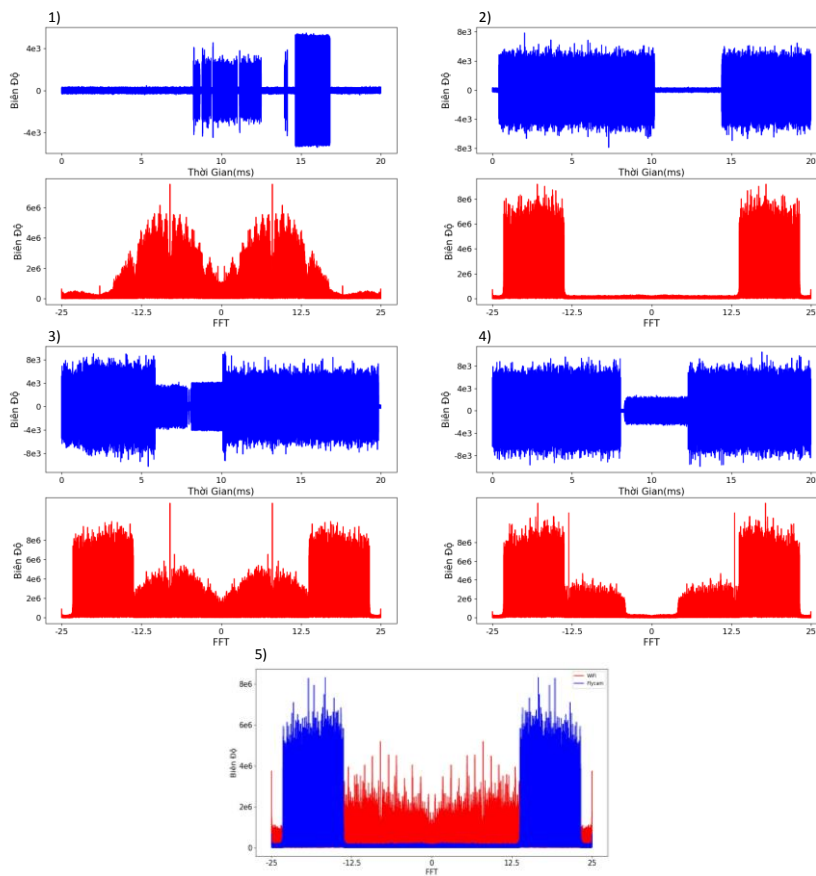
Để đánh giá hiệu quả của thuật toán lấy mẫu nén và mô hình học máy, bài báo tiến hành xây dựng 2 trường hợp thu mẫu trong hộp cách ly vô tuyến và ngoài không gian tự do, sau đó, các tập dữ liệu mô phỏng được hình thành dựa trên tập mẫu thu được từ trong hộp cách ly vô tuyến. Tập dữ liệu mô phỏng bao gồm một tập dữ liệu sử dụng nhiễu AWGN với các mức công suất khác nhau để làm giàu dữ liệu và một tập dữ liệu dùng để đánh giá sự ảnh hưởng của tín hiệu Flycam lên tín hiệu WiFi với tỉ lệ chồng lấn khác nhau ở miền tần số khi coi tín hiệu Flycam là nhiễu đối với tín hiệu WiFi.

Bộ thu USRP B210 được cài đặt ở chế độ thu với băng thông 25 MHz tần số bắt đầu là 2.400 GHz và tần số kết thúc là 2.425 GHz. Router WiFi (AP) được cài đặt phát cố định ở kênh 1 tần số trung tâm là 2.412 GHz, tần số bắt đầu là 2.401 GHz, tần số kết thúc là 2.423 GHz. Do giới hạn về phần cứng của thiết bị thu mẫu nên việc thử nghiệm chỉ tiến hành với router WiFi được cấu hình hoạt động với băng thông là 20 MHz. Bên cạnh đó, router hoạt động với băng tần 20 MHz hay 40 MHz không ảnh hưởng tới kỹ thuật phân tích tín hiệu do đều có cấu trúc tín hiệu đồng bộ preamble giống nhau. Flycam cụ thể là Flycam Phantom 3 pro được cài đặt truyền tín hiệu video ở kênh 13, tần số trung tâm là 2.406 GHz, tần số bắt đầu là 2.401 GHz và tần số kết thúc là 2.411 GHz. Anten thu được sử dụng trong bài báo là anten định hướng có dải tần từ 100 MHz-6 GHz. Bài báo tiến hành thu mẫu tín hiệu trong hộp chắn tín hiệu vô tuyến và ngoài không gian tự do được nêu ở phần tiếp theo của bài báo.

#### **A Thu tín hiệu trong hộp cách ly tín hiệu vô tuyến**

Quá trình thu mẫu được tiến hành trong hộp cách ly tín hiệu vô tuyến để đảm bảo rằng môi trường thu bên trong hộp không bị ảnh hưởng bởi các nguồn nhiễu khác từ bên ngoài mà chỉ có các tín hiệu cần khảo sát được đưa vào các đầu vào của hộp. Ngoài ra, nếu chỉ có router WiFi hoạt động độc lập thì tín hiệu thu được sẽ chỉ có bản tin broadcast nên trong mô hình sử dụng

một điện thoại di động kết nối với router WiFi để liên tục truy nhập internet được đặt trong hộp chắn tín hiệu vô tuyến nhằm tăng lưu lượng. Tín hiệu được thu trong hộp cách ly vô tuyến bao gồm các kịch bản của tín hiệu đưa vào các đầu vào của hộp là: tín hiệu WiFi độc lập, tín hiệu Flycam độc lập và tín hiệu WiFi và Flycam chồng lấn về tần số. Tập dữ liệu thu trong hộp đóng vai trò là tập dữ liệu chuẩn để đưa vào huấn luyện trong mô hình học máy đồng thời tạo ra một bộ dữ liệu mô phỏng bằng việc tổ hợp 2 loại tín hiệu WiFi và Flycam với các mức SNR khác nhau, trong bộ dữ liệu mô phỏng này, tín hiệu của Flycam là nhiễu đối với tín hiệu WiFi. Kết quả thu được được minh họa trên hình 5.



**Hình 5.** Tín hiệu thu trong hộp cách ly vô tuyến.

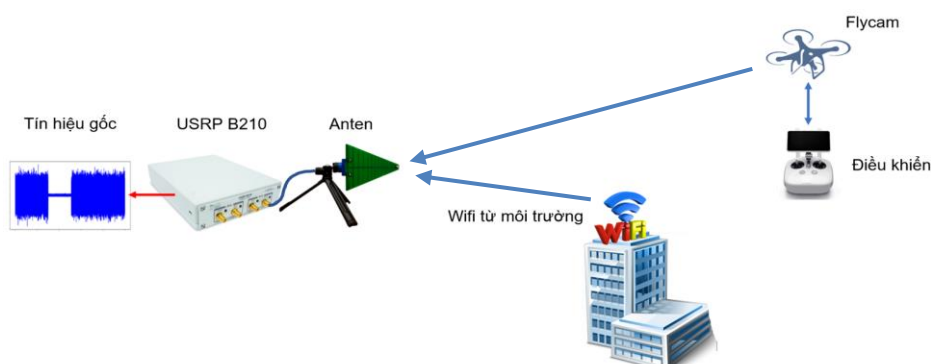
- 1) Tín hiệu WiFi độc lập. 2) Tín hiệu Flycam độc lập. 3, 4) Tín hiệu WiFi và Flycam với kênh chồng lấn. 5) Mô phỏng tín hiệu Flycam chồng lấn WiFi.

### **B Thu tín hiệu ngoài không gian tự do**

Quá trình thu mẫu được tiến hành ngoài môi trường, hệ thống được bố trí như trong hình 6.

Tín hiệu thu được bao gồm tín hiệu từ một Flycam bay ở khoảng cách 300 m so với nguồn tín hiệu WiFi được xét cùng băng tần đặt cách hệ thống thu chặn 30 m. Việc lựa chọn khoảng cách này nhằm mục đích tạo bộ tín hiệu có sự chồng lấn trong miền thời gian và miền tần số, do khoảng cách giữa Flycam và router WiFi nằm cách xa nhau nên 2 thiết bị có thể hoạt động tại cùng một băng tần mà không ảnh hưởng lẫn nhau.

Tiến hành thu 90 triệu mẫu, trong đó gồm 30 triệu mẫu WiFi và 30 triệu mẫu Flycam được thu độc lập, 30 triệu mẫu Flycam và WiFi hoạt động với băng tần chồng lấn nhau cho từng trường hợp, tổng dữ liệu thu về 180 triệu mẫu cho 2 trường hợp. Tập dữ liệu mẫu được lấy và lưu trữ như trong [14].



**Hình 6.** Mô hình hệ thống thu mẫu thực tế.

#### 4. KẾT QUẢ VÀ ĐÁNH GIÁ

Trong phần này, nội dung của bài báo đánh giá độ chính xác của mô hình và so sánh thời gian tính toán khi áp dụng lấy mẫu nén trước khi đưa vào mô hình học máy so với phương pháp được trình bày trong [8]. Các chỉ số liên quan đến nhận dạng đánh giá chính trong phương pháp LOF là 4 chỉ số precision, accuracy, recall, F1-Score được tính theo công thức (5) được tham khảo trong [8] với:

$$\begin{aligned}
 precision &= \frac{TP}{TP + FP} \\
 accuracy &= \frac{TP + TN}{TP + TN + FP + FN} \\
 recall &= \frac{TP}{TP + FN} \\
 F_1 - score &= \frac{2TP}{2TP + FP + FN}
 \end{aligned} \tag{5}$$

Trong đó, TP, TN, FP và FN lần lượt là:

True positive (TP): Là số các dữ liệu WiFi được hệ thống phân loại là WiFi, gọi là dương tính thật.

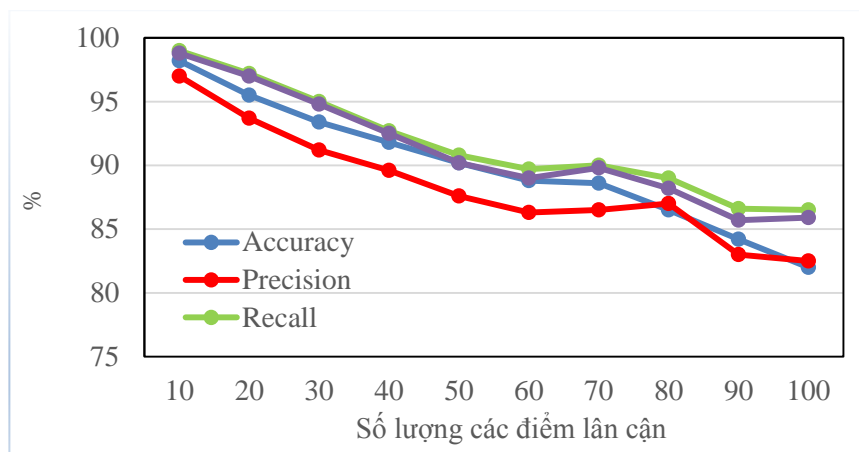
True Negative (TN): Là số các dữ liệu Flycam hệ thống phân loại là Flycam, gọi là âm tính thật.

False positive (FP): Là số các điểm dữ liệu WiFi nhưng được hệ thống đưa ra cảnh báo là tín hiệu Flycam, gọi là dương tính giả.

False Negative (FN): Là số các điểm dữ liệu là tín hiệu Flycam nhưng được hệ thống đưa ra cảnh báo là tín hiệu WiFi, gọi là âm tính giả.

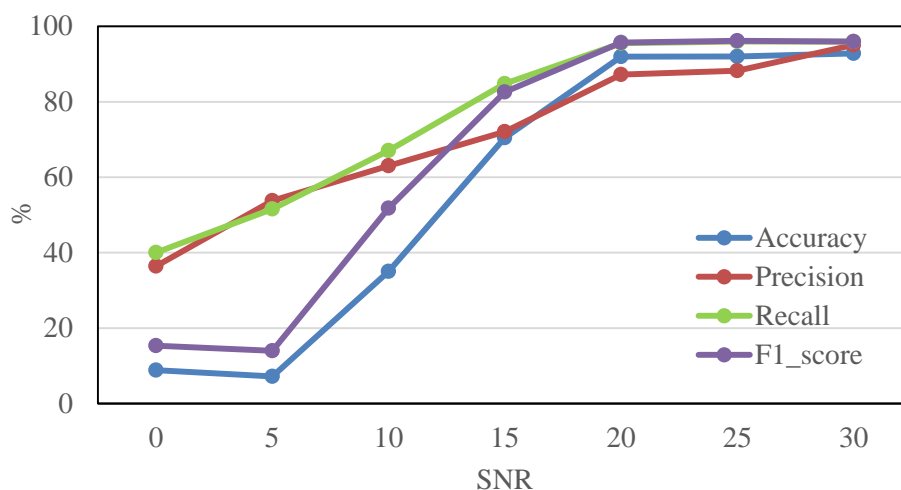
Thông số accuracy cho biết tỷ lệ số tín hiệu mô hình phân loại đúng so với toàn bộ tập dữ liệu, precision cho biết tỷ lệ số tín hiệu là WiFi so với những tín hiệu được dự đoán là WiFi, recall cho biết tỷ lệ trong số tín hiệu là WiFi có bao nhiêu tín hiệu được dự đoán đúng bởi mô hình, F1- score cho biết tỷ lệ trung bình điều hòa của 2 thông số precision và recall.

Một trong những tham số ảnh hưởng nhiều nhất đến mô hình LOF là số lượng điểm lân cận gần nhất vì nó ảnh hưởng đến hiệu suất của mô hình. Vậy nên việc chọn số lượng điểm lân cận gần nhất được thử nghiệm theo hướng tăng dần số điểm lân cận, ban đầu mô hình được đặt số lượng điểm lân cận ở mức là 10 và tăng dần đến 100, mỗi lần tăng 10 đơn vị.



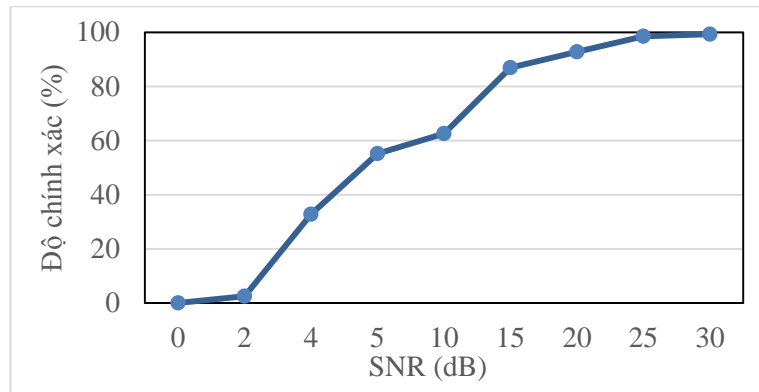
**Hình 7.** Biểu đồ khảo sát độ chính xác (%) các chỉ số đánh giá precision, recall, F1-score, accuracy của từng trường hợp số lượng điểm lân cận khác nhau.

Kết quả trong hình 7 cho thấy mỗi lần tăng số lượng điểm lân cận, độ chính xác của mô hình dựa trên bộ dữ liệu xác thực có xu hướng giảm dần và chỉ khi số lượng điểm lân cận đạt 10 thì độ chính xác của mô hình đạt cao nhất. Vì vậy, bài báo quyết định chọn số lượng điểm lân cận gần nhất bằng 10 trong mô hình LOF để đạt hiệu quả tối ưu. Tỉ số tín hiệu trên tạp âm (SNR) ở đây được định nghĩa là tỉ số của tổng công suất tín hiệu (bao gồm cả tín hiệu của Flycam và tín hiệu WiFi) trên công suất nhiễu nền. Khi khảo sát sự ảnh hưởng của số lượng điểm lân cận đến độ chính xác của mô hình, với SNR được chọn bằng 30 dB và với số lượng điểm lân cận bằng 10, các chỉ số precision, accuracy, recall, F1 score lần lượt đạt 91.26%, 93.2%, 92.43%, 91.9%. Ngoài ra, sự ảnh hưởng của nhiễu lên mô hình LOF được khảo sát bằng cách thay đổi SNR, theo đó, nhiễu trắng Gaussian thích nghi (AWGN) được thêm vào các tín hiệu thu được để có tín hiệu với SNR khác nhau, kết quả được trình bày trong hình 9.



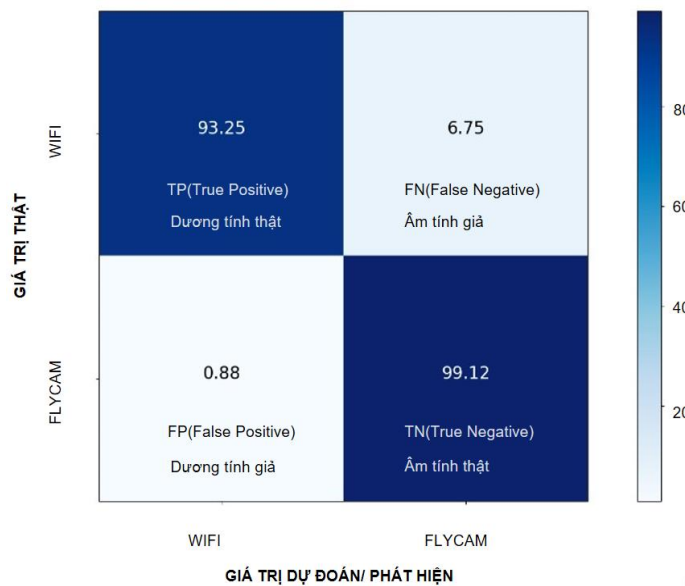
**Hình 8.** Biểu đồ biểu diễn các thông số accuracy, precision, recall, F1 score của mô hình với tập dữ liệu có SNR giảm dần.

Kết quả trong hình 8 cho thấy, khi SNR giảm thì độ chính xác của mô hình có xu hướng giảm xuống. Độ chính xác đạt đỉnh 93.25% khi SNR bằng 30 dB và duy trì ổn định dưới giá trị này khi SNR giảm xuống còn 20 dB. Độ chính xác giảm xuống còn trên 70% tại SNR bằng 15 dB. Ngoài ra, khi SNR tiếp tục giảm, số liệu về độ chính xác giảm đáng kể xuống thấp hơn 40% và 15% tương ứng khi SNR thấp hơn 10 dB và 5 dB.



**Hình 9.** Biểu đồ biểu diễn độ chính xác của mô hình trong trường hợp tín hiệu Flycam là nhiễu.

Trong trường hợp tín hiệu Flycam chồng lấn với WiFi trong miền tần số thì tín hiệu của Flycam là nhiễu đối với tín hiệu WiFi. Kết quả trong hình 9 cho thấy độ chính xác đạt mức lớn nhất lên tới 99.3% khi SNR bằng 30 dB, khi tín hiệu Flycam càng mạnh (SNR giảm) thì độ chính xác của mô hình có xu hướng giảm dần. Với mức SNR lớn hơn 15 dB, mô hình vẫn đạt hiệu quả lên tới hơn 80%. Độ chính xác giảm xuống còn dưới 62% khi SNR dưới 10 dB.



**Hình 10.** Ma trận lỗi của mô hình LOF ( $k = 10$ ) khi thực hiện tiên đoán bộ dữ liệu test với  $SNR > 25$  dB.

Giúp đánh giá kỹ hơn về mô hình, bài báo sử dụng một ma trận lỗi (confusion matrix) kích thước 2x2, trong đó, các hàng thể hiện giá trị thật, các cột thể hiện giá trị dự đoán. Kết quả trên hình 10, khi thực hiện tiên đoán bộ dữ liệu test với  $k = 10$ ,  $SNR > 25$  dB thể hiện trên hình 10 cho thấy khả năng hoạt động rất hiệu quả.

Trong [8] Ezuma và cộng sự đã trình bày phương pháp SDAE để nén dữ liệu vào đưa vào mô hình LOF nhằm phân loại hai tín hiệu có phải từ Flycam hay không cho kết quả độ chính xác là 95.7% là tín hiệu không phải Flycam và 83.3% là tín hiệu Flycam với  $SNR > 25$  dB. Nhược điểm của phương pháp này là phải tính trên toàn bộ các điểm dữ liệu được cho là có tín hiệu. Để so sánh giữa hai phương pháp này tác giả thực hiện cả hai phương pháp trên cùng tập dữ liệu của bài báo thu được với  $SNR > 25$  dB. Kết quả được trình bày ở bảng 1.

**Bảng 1.** So sánh mô hình đề xuất với mô hình trong [8] với tín hiệu có SNR > 25 dB.

	Thời gian sử dụng phương pháp mẫu nén/LOF (ms)	Thời gian sử dụng phương pháp SDAE/LOF (ms)	Độ chính xác sử dụng phương pháp mẫu nén/LOF (%)	Độ chính xác sử dụng phương pháp SDAE/LOF (%)
Lần 1	15.07	687.59	98.18	94.9
Lần 2	9.95	671.94	100	92.46
Lần 3	15.03	734.45	94.59	91.75
Lần 4	25.04	717.56	93.59	89.71
Lần 5	9.895	703.22	96.29	90.71
Lần 6	15.03	698.03	100	89.68
Lần 7	10.02	703.19	96.26	90.97
Lần 8	10.26	703.23	96.29	91.3
Lần 9	10.28	687.57	100	90.27
Lần 10	9.975	687.54	93.25	91.88
<b>Trung bình</b>	<b>13.055</b>	<b>699.432</b>	<b>96.845</b>	<b>91.363</b>

Thời gian tính toán của một mô hình LOF tăng theo cấp số nhân với số chiều của dữ liệu, đối với dữ liệu nhiều chiều độ phức tạp tính toán được biểu diễn là  $O(n^2)$  được trình bày trong [13]. Với phương pháp trong [8] mô hình LOF có độ phức tạp tính toán là  $(M/1024) * O(n^2)$  còn đối với mô hình đề xuất có độ phức tạp là  $4 * O(n^2)$ , trong đó, M là độ dài đoạn dữ liệu thu được, thực tế  $M/1024 \gg 4$ . Do đó, bằng kỹ thuật lấy mẫu nén, số chiều của dữ liệu giảm xuống đáng kể và chỉ giữ những đặc trưng của tín hiệu dự đoán. Kết quả trong bảng 1 cho thấy, tốc độ tính toán được cải thiện đáng kể do không phải tính toán trên toàn bộ các điểm dữ liệu được cho là có tín hiệu, thời gian dự đoán của mô hình giảm trung bình từ 699.4 ms xuống còn 13.05 ms, độ chính xác của phương pháp được đề xuất trong bài báo cho kết quả tốt hơn trong thực nghiệm này.

## 5. KẾT LUẬN

Bài báo này tập trung vào vấn đề phát hiện tín hiệu WiFi lẫn trong tín hiệu vô tuyến phát ra từ Flycam. Cụ thể là, trong kịch bản tín hiệu Flycam trong trường hợp phổ tần của tín hiệu thu được từ Flycam và từ các thiết bị sử dụng sóng WiFi bị chồng lấn lên nhau tại điểm thu tín hiệu, nghiên cứu đã tiến hành lấy mẫu thưa và trích chọn được 64 đặc trưng từ ma trận nén, sau đó huấn luyện trong mô hình học máy LOF để phân loại. Kết quả thử nghiệm cho thấy phương pháp sử dụng có hiệu quả hơn trong việc phát hiện tín hiệu Flycam trong môi trường có nhiều WiFi. Kết quả thử nghiệm cho thấy, độ chính xác khi tách tín hiệu WiFi trung bình lên tới hơn 93% khi SNR duy trì ở mức 20 dB đến 30 dB. Số liệu về độ chính xác còn trên 70% khi SNR bằng 15 dB và dưới 40% khi SNR thấp hơn 10 dB. Trong trường hợp tín hiệu Flycam là nhiễu của tín hiệu WiFi, mô hình đạt hiệu quả lên tới hơn 80% khi SNR lớn hơn 15 dB và dưới 62% khi SNR dưới 10 dB. Với cùng một tập dữ liệu, kết quả khi áp dụng phương pháp lấy mẫu nén so với phương pháp trong [8] cho thấy rằng cả 2 đều có độ chính xác cao, tuy nhiên, thời gian tính toán của phương pháp lấy mẫu nén nhanh hơn xấp xỉ 53 lần so với phương pháp trong [8]. Kết quả trong bài báo có thể được áp dụng trong khâu tiền xử lý để phát hiện, trước khi đưa vào phân loại và nhận dạng Flycam nhằm tối giản tính toán cho hệ thống phân loại và nhận dạng. Trong tương lai, tác giả sẽ tập trung nghiên cứu để nâng cao hiệu quả phát hiện và giảm lỗi của kỹ thuật phát hiện phụ thuộc vào SNR.

## TÀI LIỆU THAM KHẢO

- [1]. Xiao, Y., & Zhang, X. "Micro-UAV detection and identification based on radio frequency signature". In 2019 6th International Conference on Systems and Informatics (ICSAI) (pp. 1056-1062). IEEE, (2019).

- 
- [2]. Ezuma, M., Erden, F., Anjinappa, C. K., Ozdemir, O., & Guvenc, I. "Detection and classification of UAVs using RF fingerprints in the presence of Wi-Fi and Bluetooth interference". IEEE Open Journal of the Communications Society, 1, pp. 60-76, (2019).
- [3]. Zuo, M., Xie, S., Zhang, X., & Yang, M.: "Recognition of UAV video signal using RF fingerprints in the presence of WiFi interference". IEEE Access, 9, 88844-88851.
- [4]. Boulogeorgos, A. A. A., Chatzidiamantis, N. D., & Karagiannidis, G. K. "Energy detection spectrum sensing under RF imperfections". IEEE Transactions on Communications, 64(7), 2754-2766, (2016).
- [5]. Schmidl, T. M., & Cox, D. C. "Robust frequency and timing synchronization for OFDM". IEEE transactions on communications, 45(12), pp. 1613-1621, (1997).
- [6]. Baraniuk R., "Compressive sensing", IEEE Sig. Proc. Mag. 24, no. 4, pp. 118-121, (2007).
- [7]. Xu, Guangwu, and Zhiqiang Xu. "Compressed sensing matrices from Fourier matrices." IEEE Transactions on Information Theory 61.1: pp. 469-478, (2014).
- [8]. Medaiyese, O. O., Ezuma, M., Lauf, A. P., & Adeniran, A. A. "Hierarchical Learning Framework for UAV Detection and Identification". IEEE Journal of Radio Frequency Identification, 6, pp. 176-188, (2022).
- [9]. Medaiyese, O. O., Syed, A., & Lauf, A. P. "Machine learning framework for RF-based drone detection and identification system". In 2021 2nd International Conference On Smart Cities, Automation & Intelligent Computing Systems (ICON-SONICS) (pp. 58-64). IEEE, (2021).
- [10]. Kılıç, R., Kumbasar, N., Oral, E. A., & Ozbek, I. Y. "Drone classification using RF signal based spectral features". Engineering Science and Technology, an International Journal, 28, 101028, (2022).
- [11]. Ezuma, M., Erden, F., Anjinappa, C. K., Ozdemir, O., & Guvenc, I. "Detection and Classification of UAVs Using RF Fingerprints in the Presence of Interference". arXiv preprint arXiv:1909.05429, (2019).
- [12]. Gogoi, P., Bhattacharyya, D. K., Borah, B., & Kalita, J. K. "A survey of outlier detection methods in network anomaly identification". The Computer Journal, 54(4), pp. 570-588, (2011).
- [13]. Breunig, M. M., Kriegel, H. P., Ng, R. T., & Sander, J. "LOF: identifying density-based local outliers". In Proceedings of the 2000 ACM SIGMOD international conference on Management of data, pp. 93-104, (2000).
- [14]. Allahham, M. S., Al-Sa'd, M. F., Al-Ali, A., Mohamed, A., Khattab, T., & Erbad, A. "DroneRF dataset: A dataset of drones for RF-based detection, classification and identification". Data in brief, 26, 104313, (2019).

### ABSTRACT

#### **A proposed model for processing UAV signals in the presence of WiFi co-channel interference based on compressive sensing and machine learning**

*This paper proposes a model for detecting WiFi signals in the presence of co-channel Flycam signals. The proposed method uses compressed sensing and machine learning techniques in the case that Flycam and WiFi signals are co-channel. A particular compressed sensing matrix is designed to sample the first period of the signal segment, including 256 samples. The matrix output is a vector 1x64 which is used as 64 features of the WiFi signal for the AI model based on the LOF algorithm to classify WiFi signals and non-WiFi signals. The number of samples required to remove WiFi from the mixed signal is reduced by extracting the features of the WiFi signal at the beginning of the signal segment rather than the entire signal sample. The WiFi signal classification method using the proposed model has an accuracy of 93.25% with SNR = 30 dB and above 70% with SNR = 15 dB and a faster execution time than formula-based and calculation-based classification with the total number of samples.*

**Keywords:** Compressive sensing; WiFi interference; Machine learning.