

딥보이스로 제작된 음성 탐지 시스템 연구

Researching speech detection systems made with DeepVoice

중부대학교 정보보호학전공

정진호, 박우경, 이하영, 최수민, 홍준희 | 유승재, 김민수(지도교수)



목차

I. 서론 및 관련 연구

1. 연구의 필요성
2. 관련 연구 자료

II. 제안하는 방법

1. 연구 방향성
2. 피처 선정
3. 머신러닝 & 딥러닝

III. 앱 개발

1. 딥보이스 음성 탐지
애플리케이션 제작
2. 시연 영상

IV. 검증 & 결과

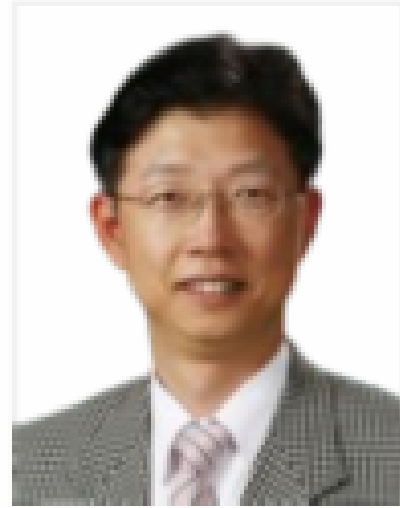
1. 연구 결과
2. 기대효과
3. 성과

초록

팀원 소개 및 프로젝트 WBS

0. 초록

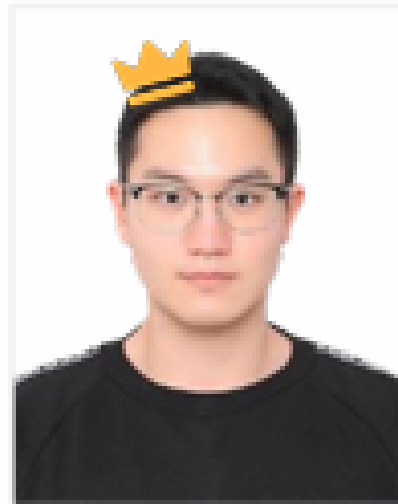
팀원 소개



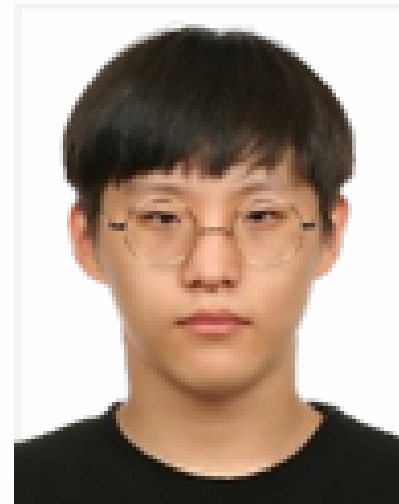
유승재 교수님



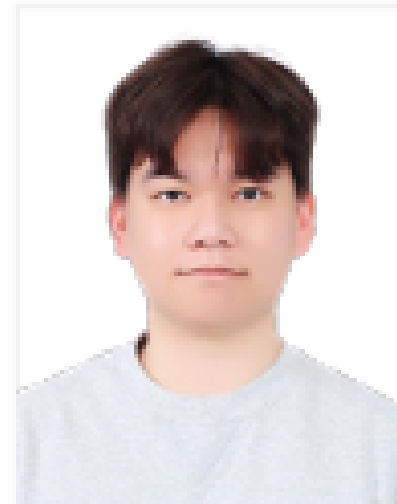
김민수 교수님



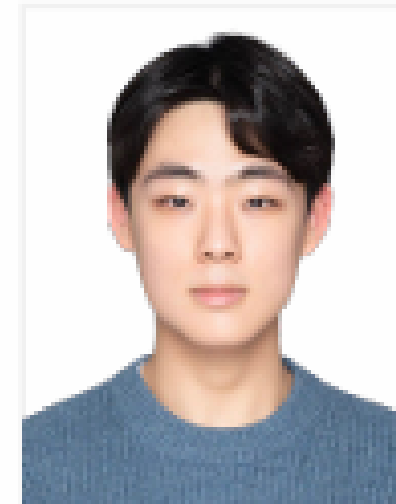
최수민



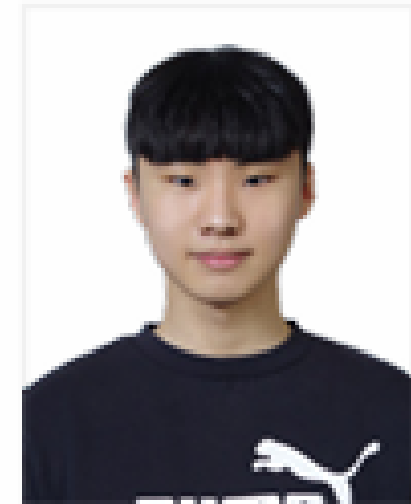
박우경



이하영



정진호



홍준희

0. 초록

WBS

	3	4	5	6	7	8	9	10	11
주제 선정 및 자료 조사									
기본 설계									
딥러닝 설계									
테스트 및 수정									
중간 발표									

서론 및 관련 연구

연구의 필요성 및 관련 연구 자료

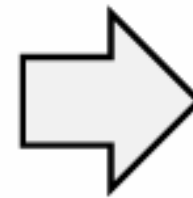
I. 서론 및 관련 연구

연구의 필요성 및 관련 연구 자료

DeepFake

Deep Learning과 Fake의 합성어

인공지능 기술을 사용하여 현실에 존재하지 않는 것이나 발생한 적이 없는 사건을 이미지, 비디오, 오디오 등을 합성



I. 서론 및 관련 연구

연구의 필요성 및 관련 연구 자료

DeepVoice

Deep Learning과 Voice의 합성어

삼성의 'Bixby', 애플의 'Siri', 구글의 "Assistant", 아마존의 "Alexa"

딥보이스 기술을 이용한 보이스피싱 피해 사례 급증, 특히 금융 피해 건수 및 금액 증가

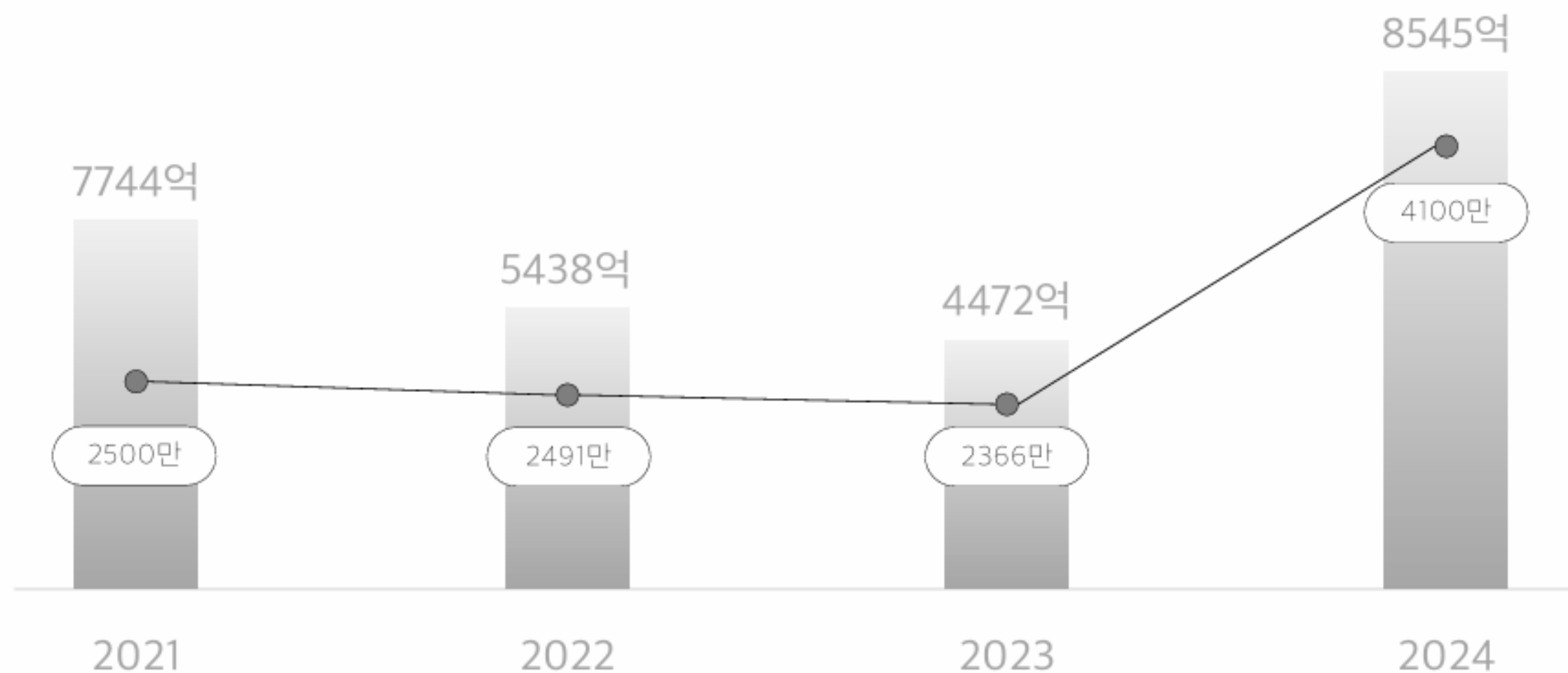


I. 서론 및 관련 연구

연구의 필요성 및 관련 연구 자료

보이스피싱 피해 현황 및 추이

● 1인당 피해액 추이 ■ 총 피해액



출처: 경찰청

I. 서론 및 관련 연구

연구의 필요성 및 관련 연구 자료

[bnTV] 내 목소리 빼앗는 ‘딥보이스 피싱’ 주의보

입력 : 2024-09-11 10:55

가

가

f

X

DM

N

상용화된 음성복제 AI 서비스, 목소리 샘플로 특정인의 어투·감정표현까지 가능해 ‘여보세요’ 한 마디만으로 목소리 복제 가능...딥보이스 피싱 피해 확대

예방법은 무반응 전화에는 무응답, 당사자들 간만 소통 가능한 ‘암호 설정’ 등

신기술 범죄 예방 기술 개발 및 플랫폼 차원의 대응 필요성 대두

출처: 보안뉴스

"딸이 납치됐어요" .. 딥페이크 영상 이용한 보이스피싱 범죄 '주의'

김인영 기자 2024.11.07 13:28:31

경찰이 최근 딥페이크 영상을 이용한 납치 범죄 보이스피싱(전화금융사기) 범죄가 발생했다며 주의를 당부했다. /이미지=이미지투데이

딥페이크 영상을 이용한 납치 범죄 보이스피싱(전화금융사기) 범죄가 발생하고 있다.

7일 뉴스1에 따르면 경찰청 국가수사본부는 이날 최근 한국을 여행 중인 딸이 납치됐다는 외국인을 대상으로 한 딥페이크 전화 금융사기가 발생했다고 밝혔다.

지난달 한 외국인은 외국 범죄 조직으로부터 한국을 여행 중인 딸이 방 안에 갇힌 채 울면서 살려달라고 하는 영상을 받았다. 범인은 합의금을 요구하며 협박했고 부모는 이 사실을 영상관에 알려 해당 사건이 한국 경찰에 신고됐다.

경찰은 딸의 안전을 확인했고 해당 영상은 딥페이크 방식의 허위 영상물인 것으로 확인됐다.

출처: 머니S

"강도·폭행 당했다고"..김성주→이형택, '구사창' 피싱 사기에 속았다 [Oh!쎈 이슈]

OSEN 김나연 기자

발행 2024.09.07 23:02

출처: OSEN

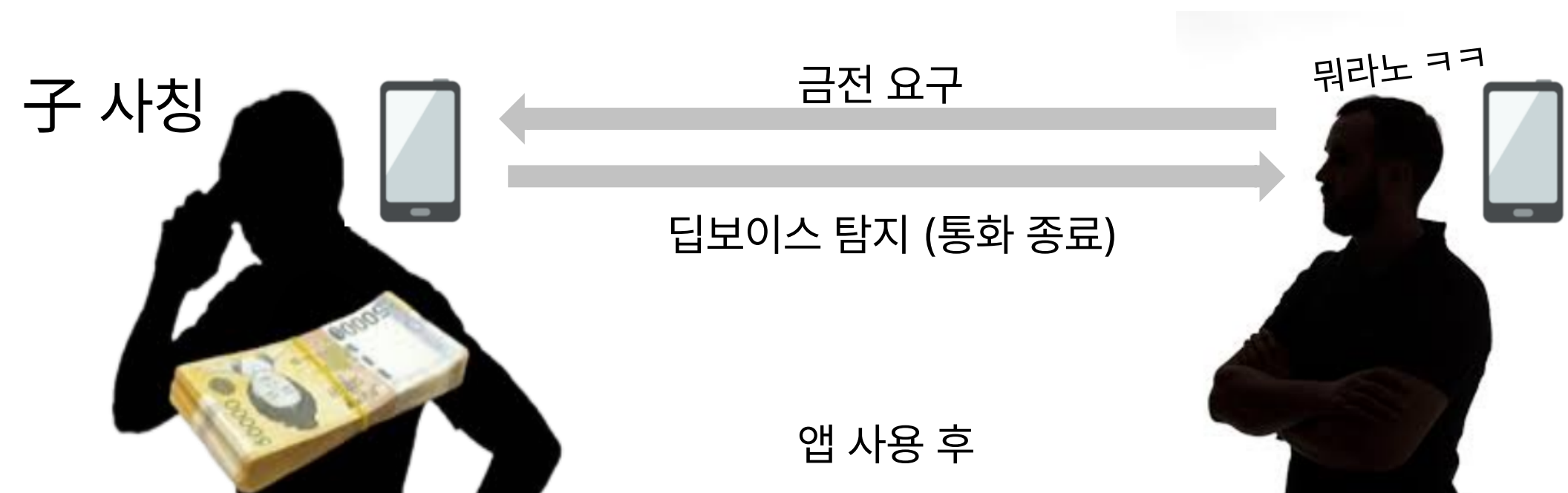
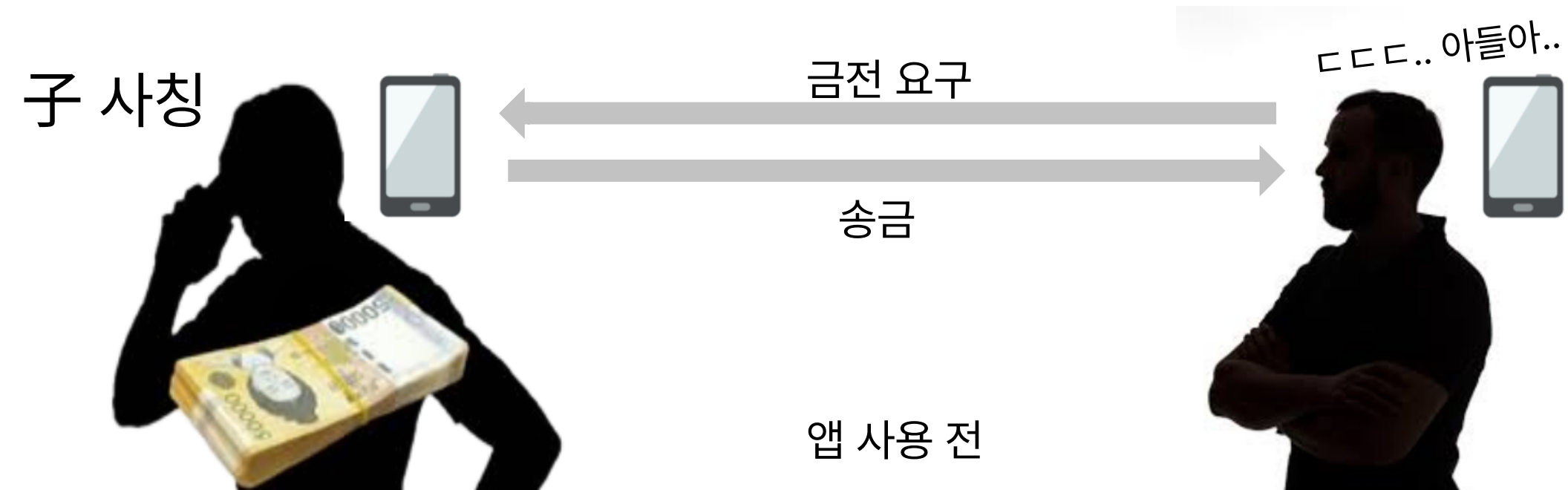
보이스 피싱 범죄 방식이 나날이 진화하고 있다. 최근에는 AI 기술의 발전으로 가족이나 지인의 목소리를 녹음하고 딥 러닝을 통해 모방, 피싱 사기에 이용하는 사례도 등장했다. 이에 더해 SNS에 올라온 사진으로 AI 아바타를 만들어 영상통화까지 하는 수준에 도달했다. 얼굴, 목소리까지 같으니 피해자는 골똥없이 속아넘어갈 수밖에 없는 상황이다. 그 결과 금융감독원 조사 결과에 따르면 지난해 보이스피싱 총 피해액은 1965억원으로, 전년보다 514억원(35.4%)나 증가한 것으로 알려졌다.

불과 몇년 전까지만해도 보이스피싱은 정보에 어두운 노인들이나 비교적 판단력이 흐린 사람이 주 피해층이라는 인식이 강했지만, 점차 교묘해지는 수법에 그 누구도 안도할수 없는 상황에 놓였다. 특히 방송을 통해 가족이 공개된 유명인의 경우 '딥페이크'를 넣은 '딥보이스'를 이용한 범죄에 노출되기 더 쉬운 상황. 실제 이들은 예능에서 가족을 사칭한 보이스 피싱에 속아넘어갔던 경험을 털어낸 경각심을 더욱 일깨운다.

DeepVoice를 이용하여 가족과 관련된 보이스피싱 문제 多

I. 서론 및 관련 연구

연구의 필요성 및 관련 연구 자료



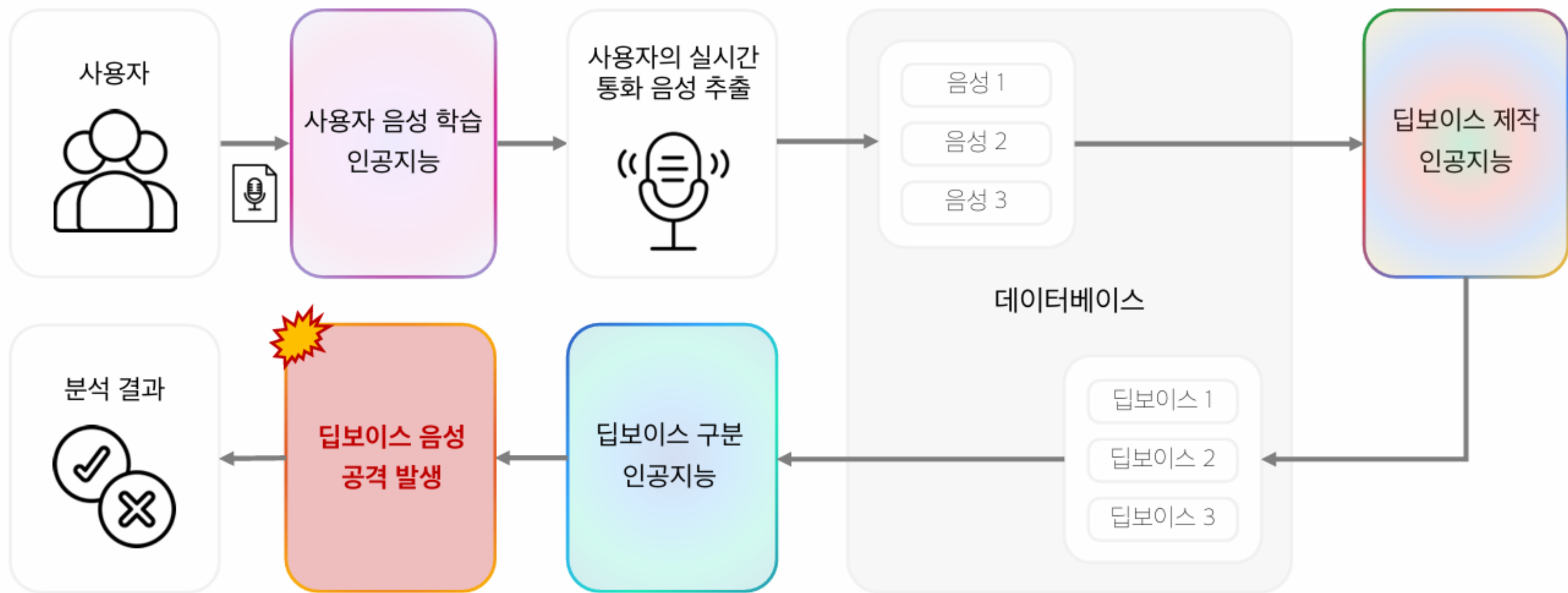
딥보이스 탐지 애플리케이션(전화 통화) 개발

제안하는 방법

피처 선정 및 머신러닝 & 딥러닝

II. 제안하는 방법

연구 방향성



II. 제안하는 방법

피처 선정 [모든 피처]

MFCC	Spectral Bandwidth	Spectral Contrast	Chroma
Tonnetz	Zero-Crossing Rate	Spectral Centroid	Spectral Rolloff
Spectral Flatness	RMS Energy	Harmonic/ Percussive	Tempo

MFCC : 60 + Chroma : 12 + Contrast : 7 + Tonnetz : 6 + ZCR : 1 + Spectral Centroid, Rolloff, Bandwidth : 3 + flatness & rms : 2 + hnr & tempo : 2 = 총 93차원

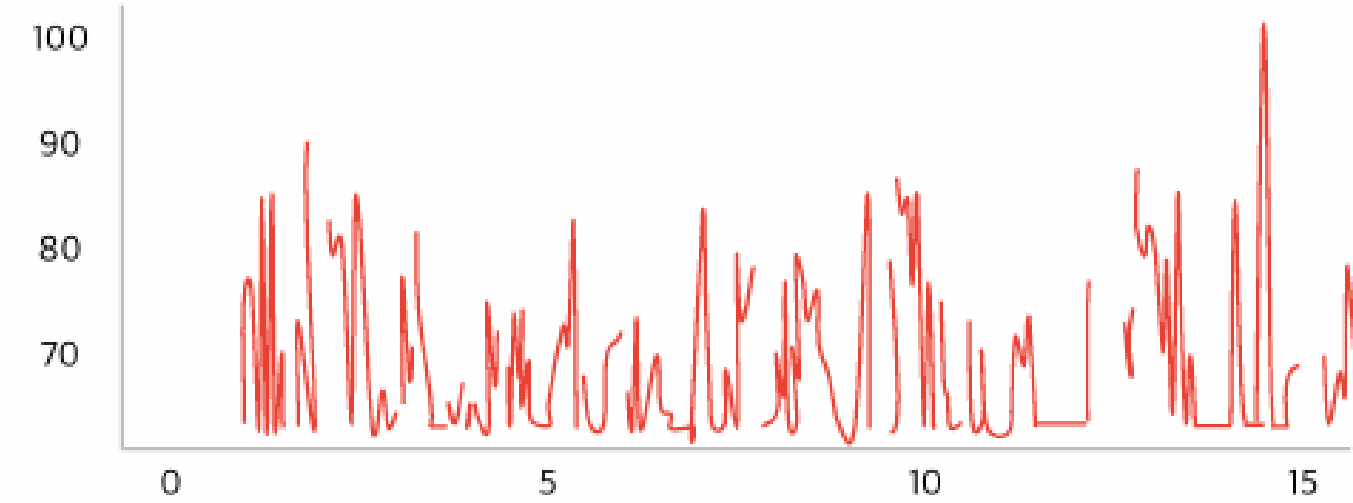
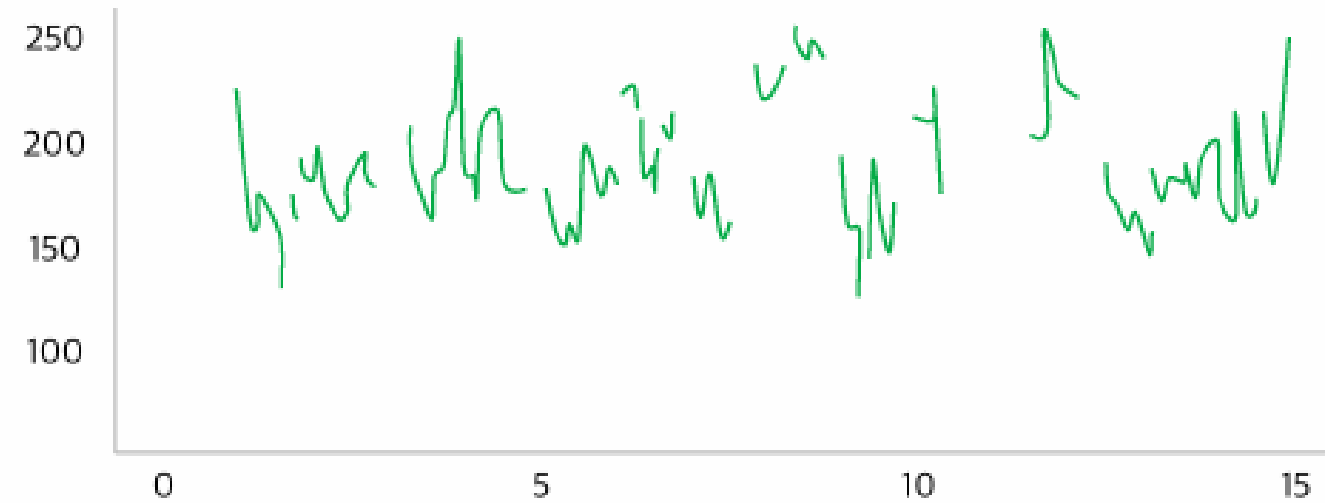
II. 제안하는 방법

피쳐 선정 01 [기본 주파수 변화]

기본 주파수 변화 [Pitch Control]

● 정상 ● 딥보이스

가로: 시간(초) | 세로: 주파수(Hz)



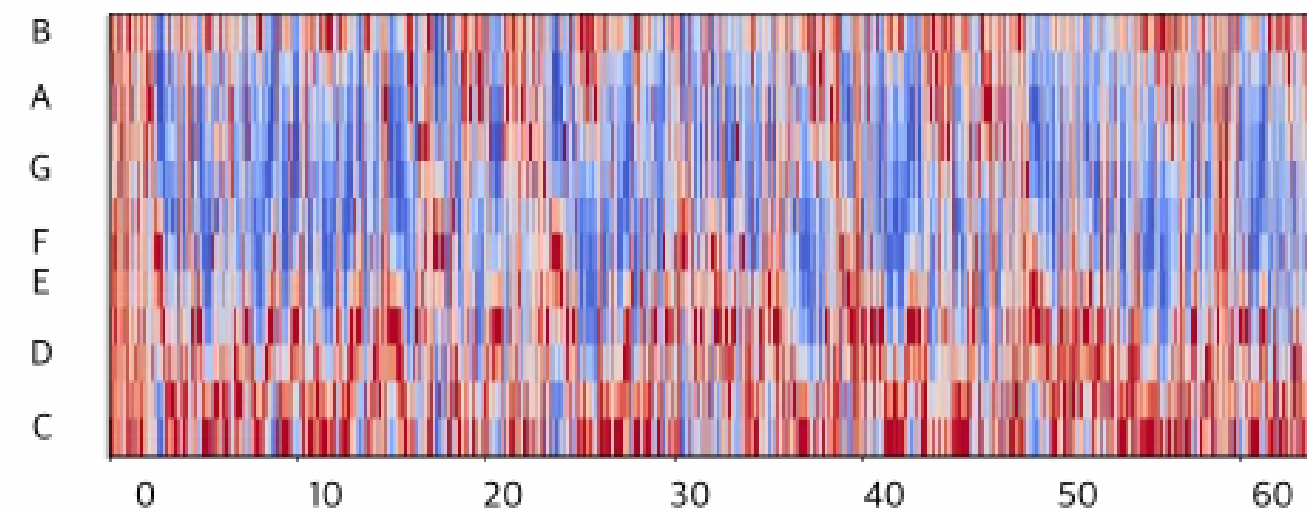
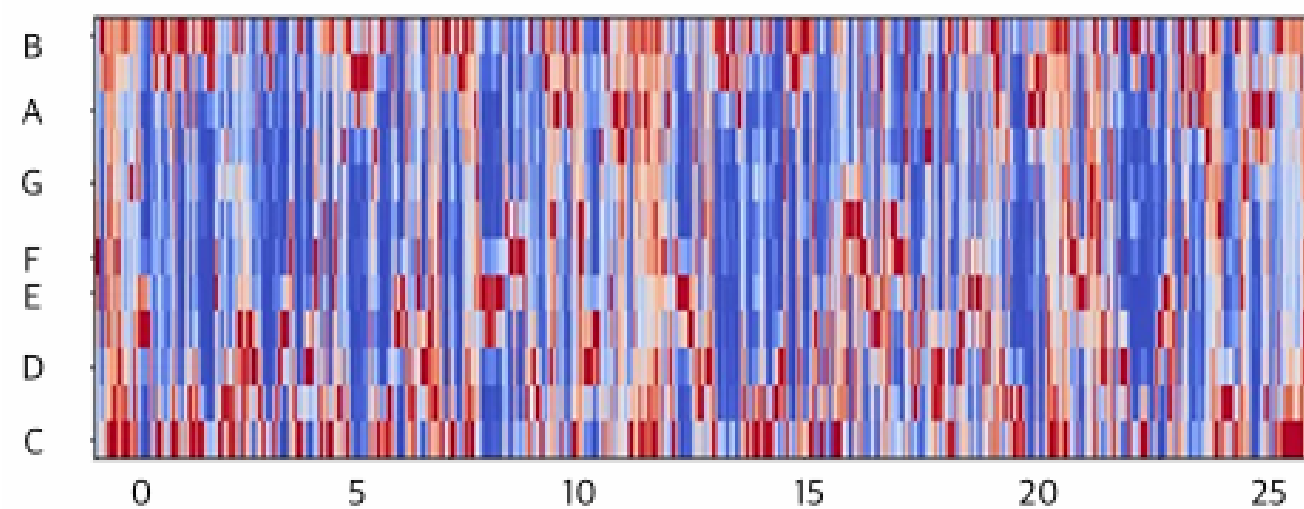
II. 제안하는 방법

피쳐 선정 02 [12 반음계 변화]

12반음계 변화 [Chroma Feature]

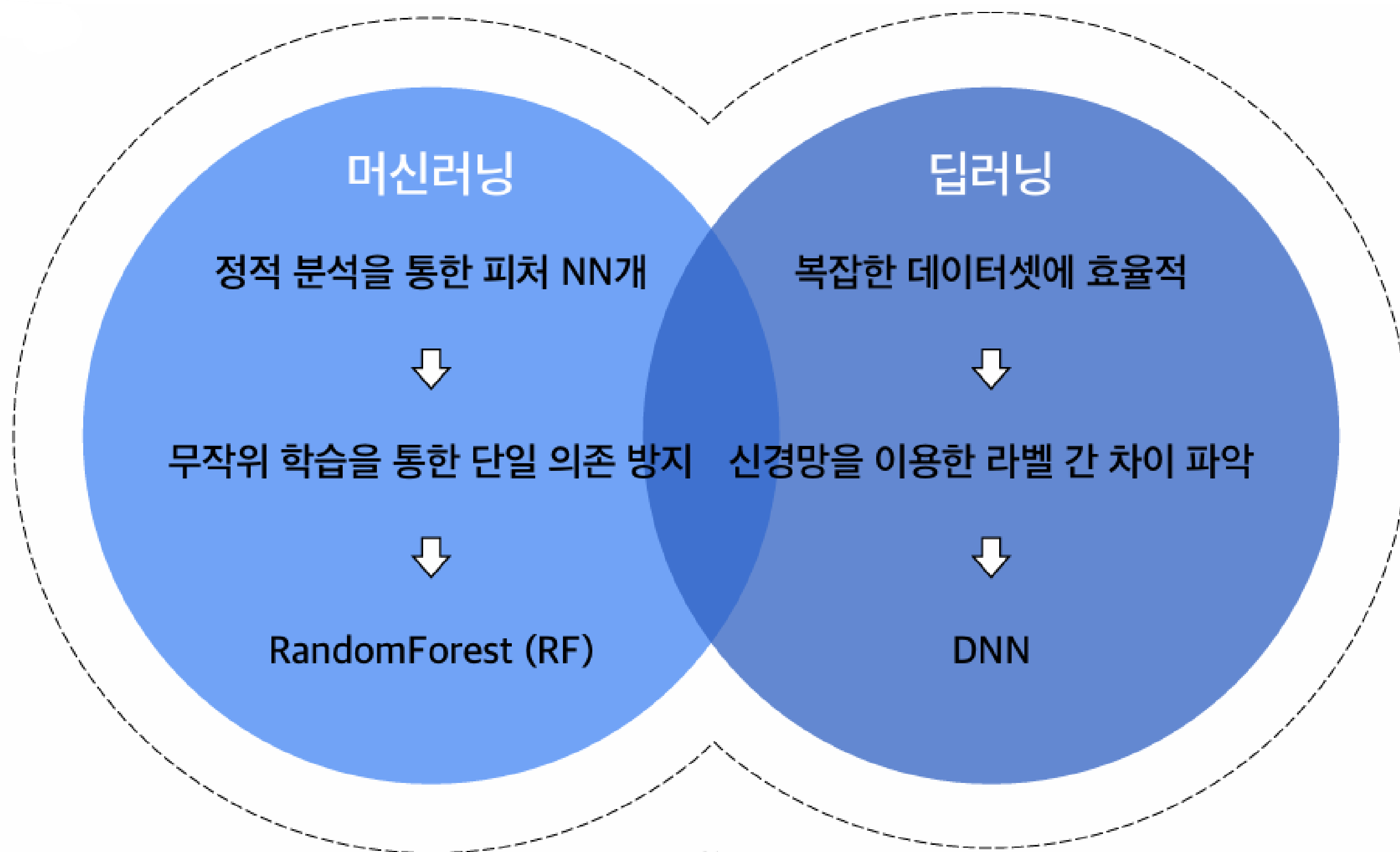
0  1

가로: 시간(초) | 세로: 12반음계



II. 제안하는 방법

머신러닝 & 딥러닝



II. 제안하는 방법

Scaler(스케일러)

MFCC	Spectral Bandwidth	Spectral Contrast	Chroma
Tonnetz	Zero-Crossing Rate	Spectral Centroid	Spectral Rolloff
Spectral Flatness	RMS Energy	Harmonic/Percussive	Tempo

MFCC : 60 + Chroma : 12 + Contrast : 7 + Tonnetz : 6 + ZCR : 1 + Spectral Centroid, Rolloff, Bandwidth : 3 + flatness & rms : 2 + hnr & tempo : 2 = 총 93차원



단위 / 크기



```
# MFCC: 20차원 + Delta + Delta-Delta (총 60차원)
mfcc = librosa.feature.mfcc(y=y, sr=sr, n_mfcc=20)
mfcc_mean = np.mean(mfcc, axis=1)
mfcc_delta = np.mean(librosa.feature.delta(mfcc), axis=1)
mfcc_delta2 = np.mean(librosa.feature.delta(mfcc, order=2), axis=1)

# Chroma: 12차원 평균
chroma = librosa.feature.chroma_stft(y=y, sr=sr)
chroma_mean = np.mean(chroma, axis=1)

# Spectral Contrast: 7차원 평균
contrast = librosa.feature.spectral_contrast(y=y, sr=sr)
contrast_mean = np.mean(contrast, axis=1)

# Tonnetz: 6차원 평균 (harmonic 부분 사용)
y_harmonic = librosa.effects.harmonic(y)
tonnetz = librosa.feature.tonnetz(y=y_harmonic, sr=sr)
tonnetz_mean = np.mean(tonnetz, axis=1)
```

```
# 피쳐 스케일링
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 모델 저장 디렉토리 생성
if not os.path.exists('./model'):
    os.makedirs('./model')
joblib.dump(scaler, './model/scaler01.pkl')
```

II. 제안하는 방법

머신러닝, RF (Random Forest)

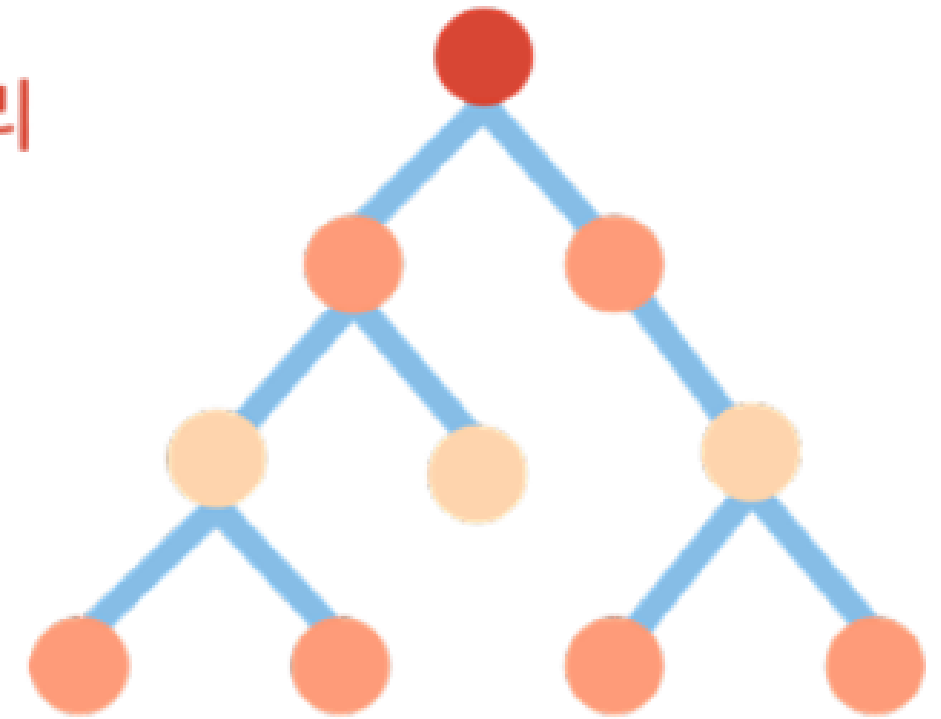
작은·중간 데이터에 강함: 배경으로 과적합 완화, 안정적 일반화

노이즈·이상치에 강건: 녹음품질 편차, 짧은 클립 등 잡음 많은 샘플에서 성능 급락이 적어 앙상블의 안전판 역할

하이퍼파라미터 민감도 낮음: 큰 튜닝 없이도 준수한 성능 → 실험 반복 시 변동폭(분산) 완화.

비선형·상호작용 즉시 포착: 91차 수공 피처(MFCC/Chroma/Contrast/Tonnetz 등) 간의 규칙형 관계를 트리 분할로 빠르게 학습.

순차적인
의사결정 트리



```
# 모델 1: Random Forest 분류기 (트리 수를 늘려 복잡한 패턴을 학습)
rf_model = RandomForestClassifier(n_estimators=200, random_state=42)
rf_model.fit(X_scaled, y)
joblib.dump(rf_model, './model/rf_model01.pkl')
```

II. 제안하는 방법

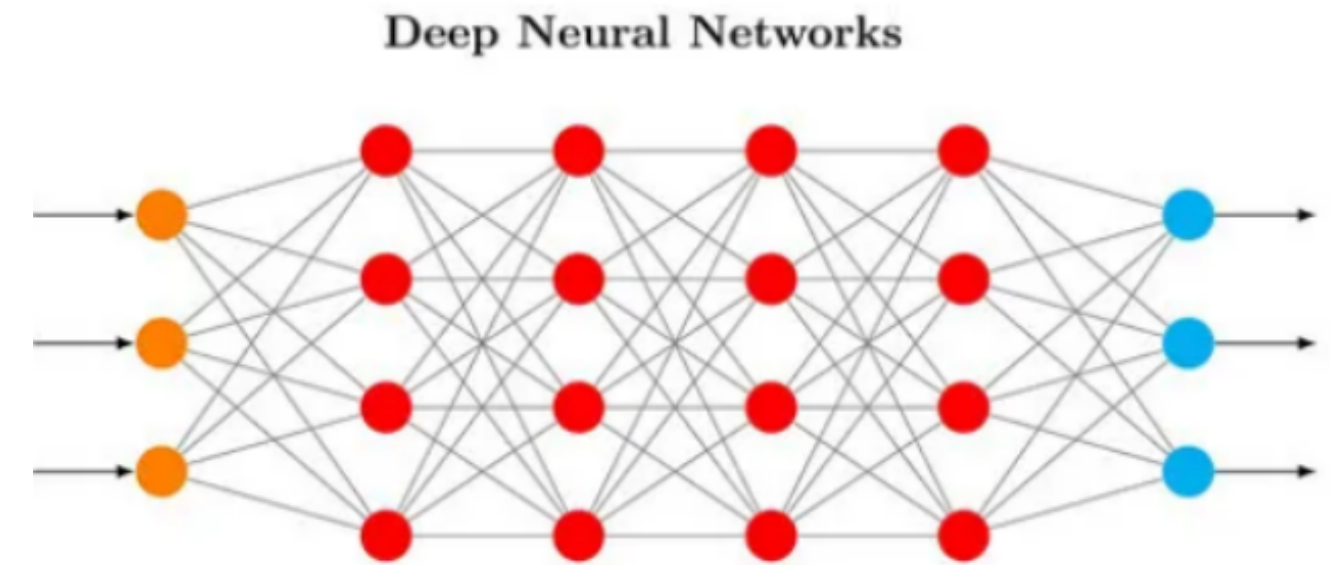
딥러닝, DNN (Deep Neural Network)

```
# 모델 2: 딥러닝 모델 (DNN) - 층 수와 드롭아웃 비율 조정
input_dim = X_scaled.shape[1]
dnn_model = Sequential()
dnn_model.add(Dense(128, activation='relu', input_dim=input_dim))
dnn_model.add(Dropout(0.4))
dnn_model.add(Dense(64, activation='relu'))
dnn_model.add(Dropout(0.4))
dnn_model.add(Dense(32, activation='relu'))
dnn_model.add(Dropout(0.4))
dnn_model.add(Dense(1, activation='sigmoid'))

dnn_model.compile(optimizer=Adam(learning_rate=0.001), loss='binary_crossentropy', metrics=['accuracy'])
early_stop = EarlyStopping(monitor='loss', patience=10, restore_best_weights=True)
dnn_model.fit(X_scaled, y, epochs=150, batch_size=8, verbose=1, callbacks=[early_stop])
dnn_model.save('./model/dnn_model01.h5')
```

```
# 각 모델의 예측 확률 (본인일 확률)
rf_probs = rf_model.predict_proba(X_test_scaled)[: , 1]
dnn_probs = dnn_model.predict(X_test_scaled).flatten()

# 두 모델의 확률 평균으로 ensemble (평균 확률 >= 0.80이면 본인(1), 아니면 타인(0))
ensemble_probs = (rf_probs + dnn_probs) / 2.0
ensemble_preds = (ensemble_probs >= 0.80).astype(int)
```



모델의 최종 확률 계산

RF = RandomForest (RF) 예측 확률

DNN = DNN 예측 확률

Total = RF + DNN = 두 모델의 확률 합계

Value = Total / 2 = 최종 모델의 확률

Value >= 80 (실제 음성), Value < 80 (딥보이스)

앱 개발

딥보이스 음성 탐지 애플리케이션 제작

III. 앱 개발

기술 스택

서버



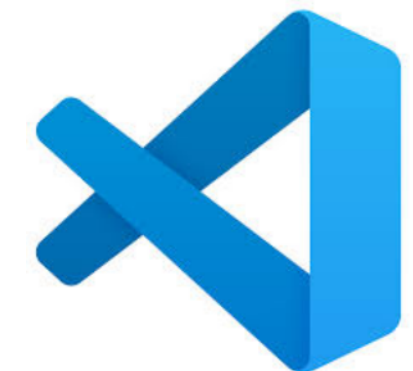
앱



플랫폼

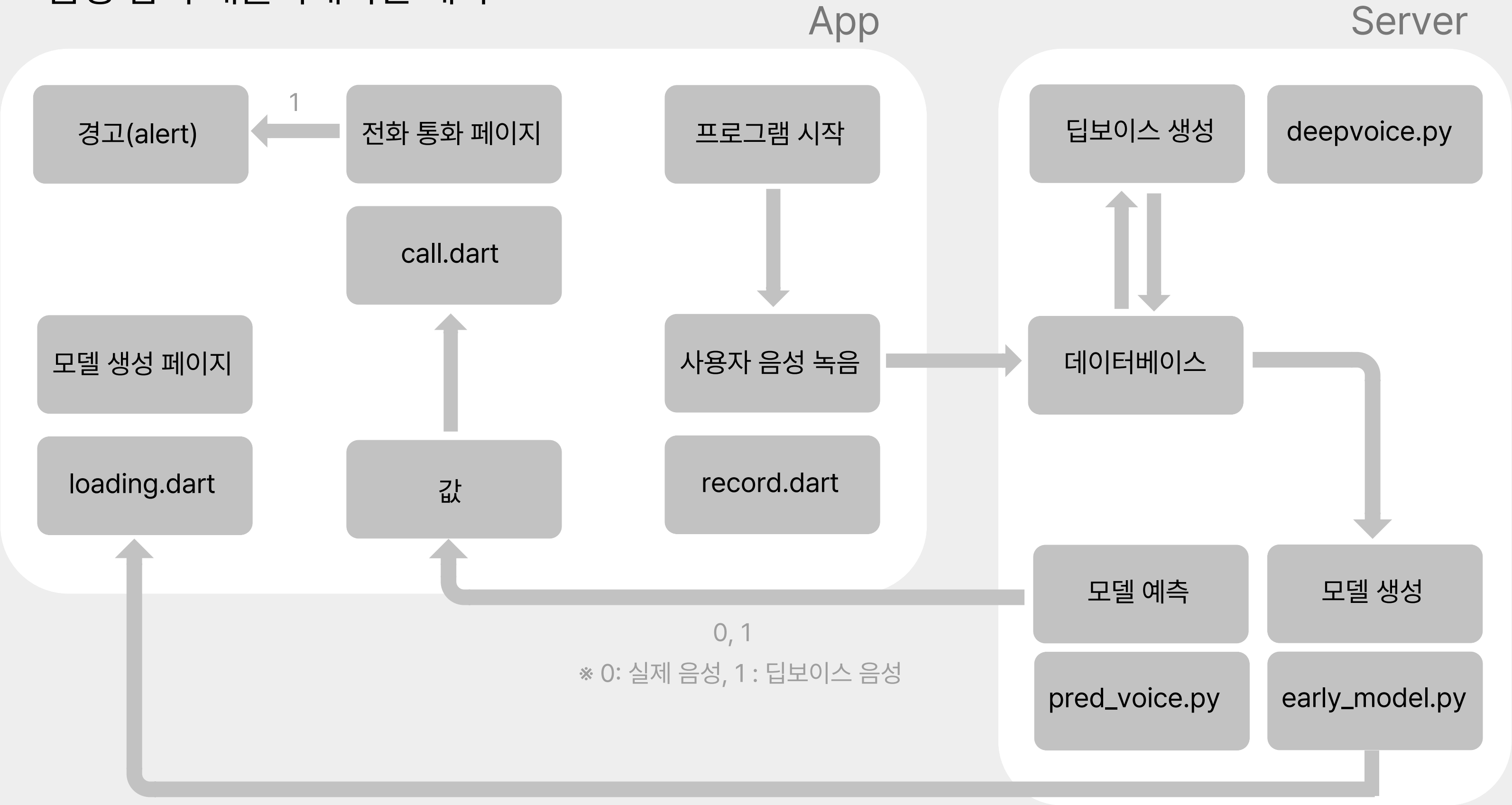


머신러닝 & 딥러닝



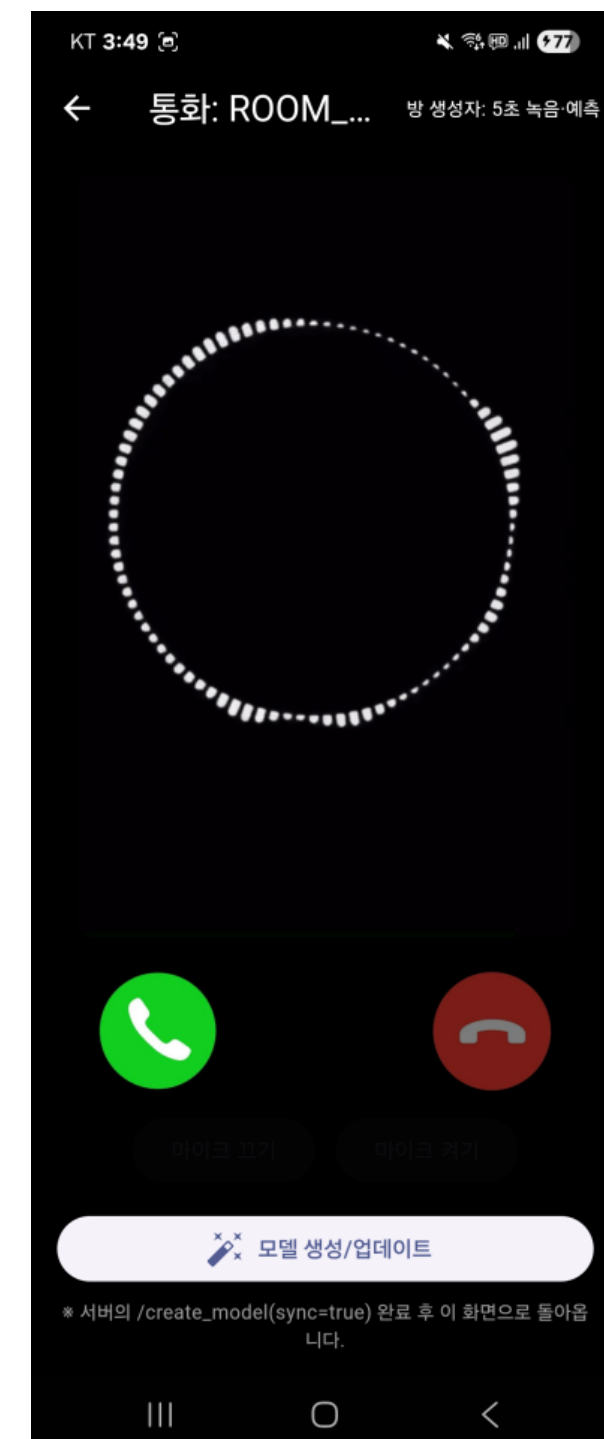
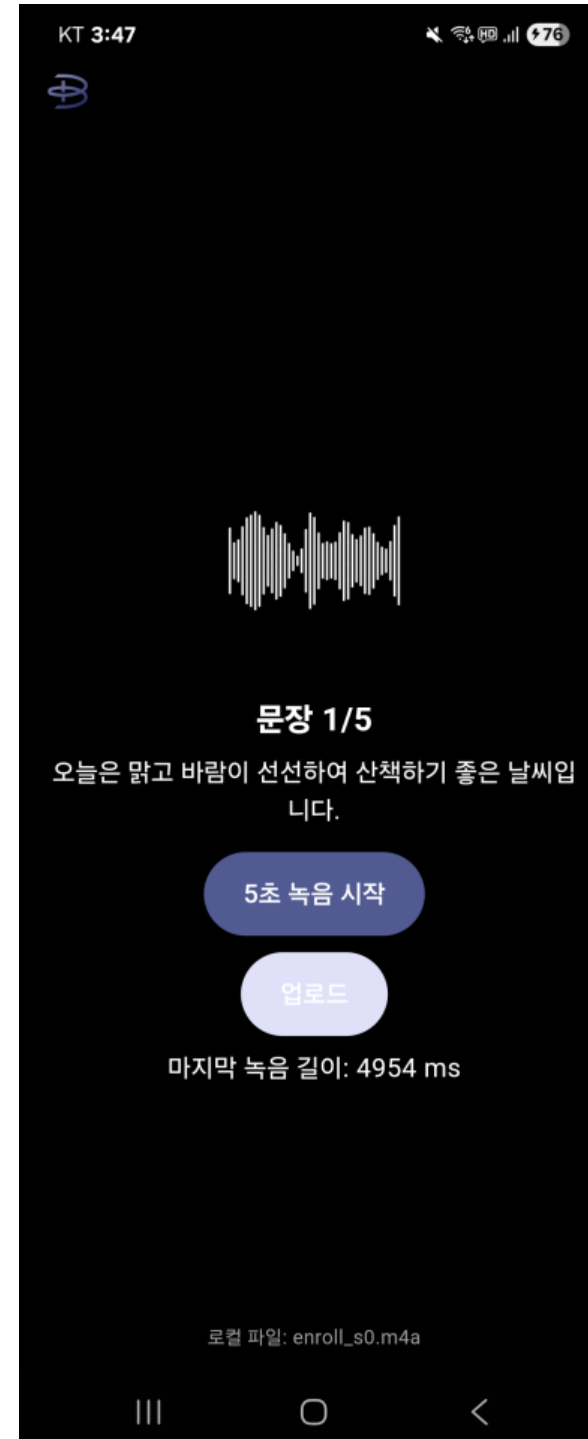
III. 앱 개발

딥보이스 음성 탐지 애플리케이션 제작



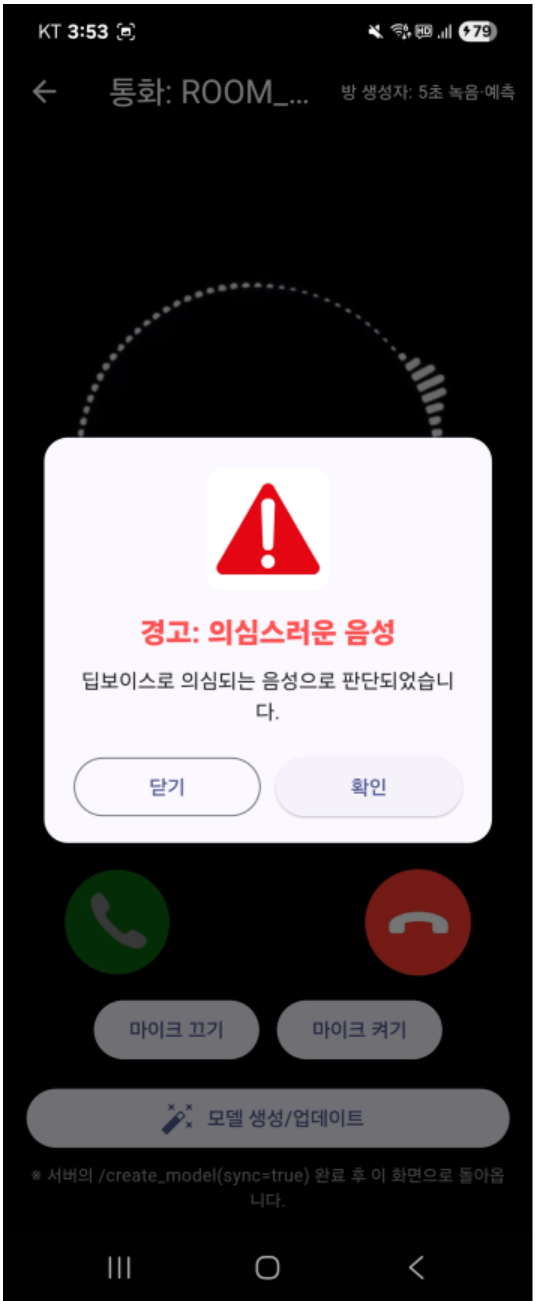
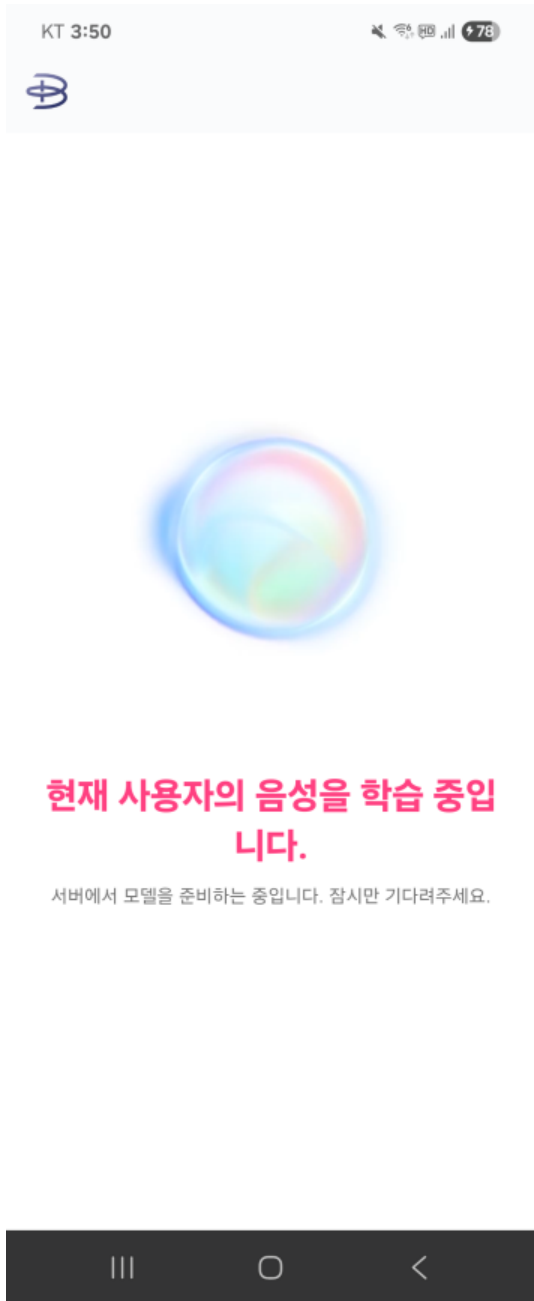
III. 앱 개발

딥보이스 음성 탐지 애플리케이션 제작

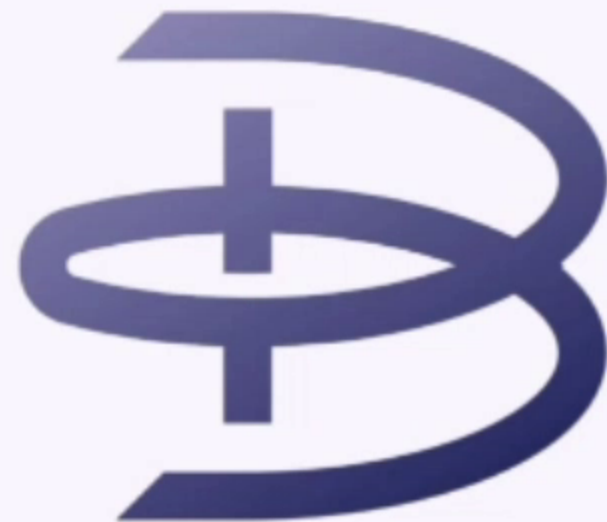


III. 앱 개발

딥보이스 음성 탐지 애플리케이션 제작



시연 영상



UUID: ANON_72f355b6

시작하기

검증 & 결과

연구결과및참고문헌

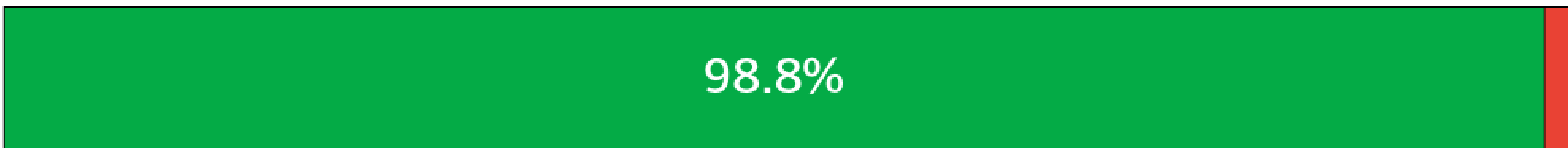
IV. 검증 & 결과

연구 결과

딥보이스 음성 탐지

● 정상 ● 오탐

테스트 크기: 992개 | 가로: 정확도 (%)



0 10 20 30 40 50 60 70 80 90 100

IV. 검증 & 결과

기대 효과

①

사이버 범죄

보이스피싱, 금융 피해 대응

음성 관련 저작권 보호

신분 도용 및 명예 훼손 방지

②

이용률에 따른 정확도 향상

시스템 운영의 효율적인 자원 활용

새로운 보안 인프라 구축

인공지능 학습 시스템 신뢰도 향상

③

기술의 폭넓은 활용

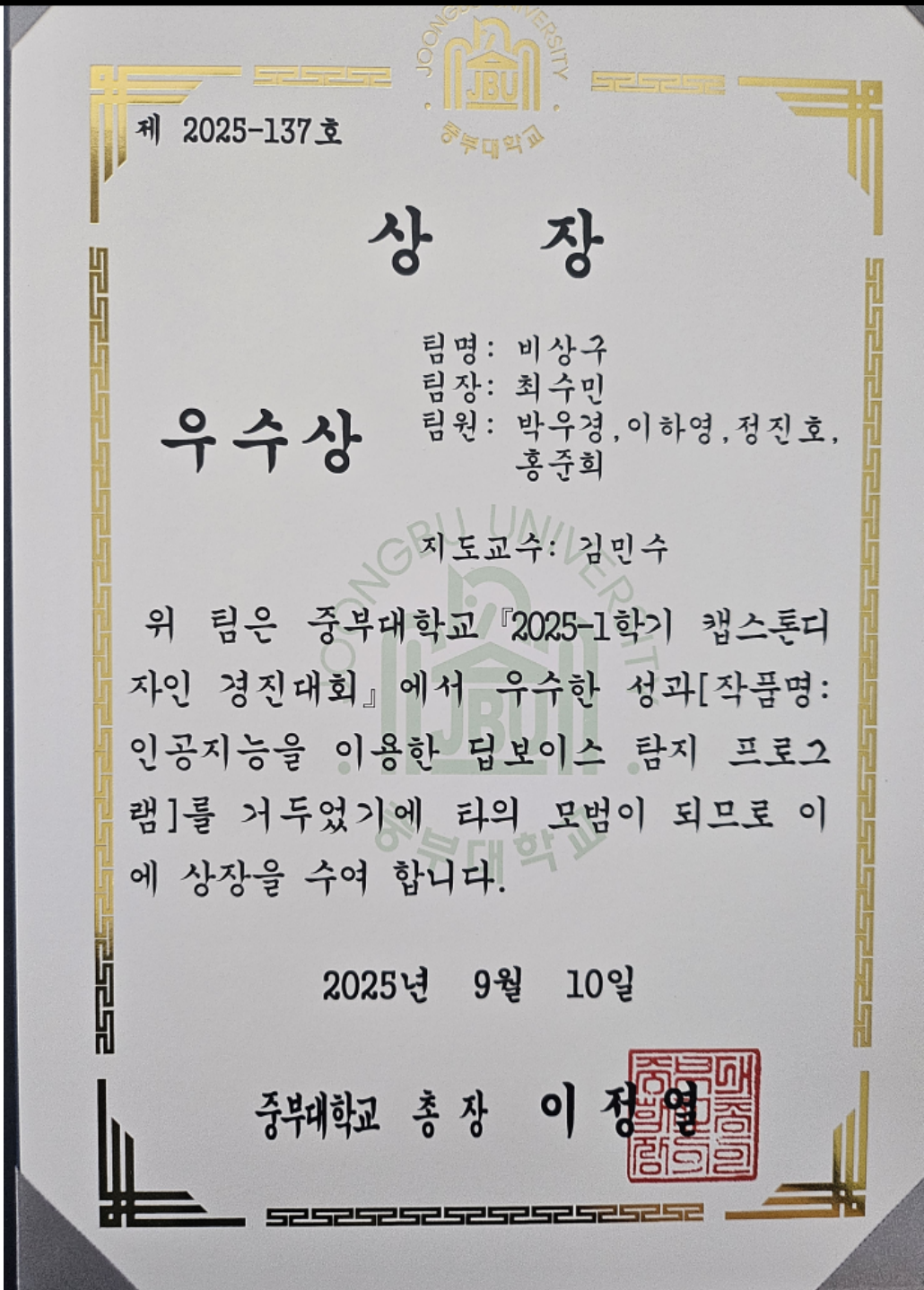
API 제공을 통한 연계 가능

음성 기반 증거 분석 활용 제공

다양한 산업 분야의 가능성

IV. 검증 & 결과

성과

 2025-30 표 창 장 우수논문상 논문 제목 : 딥보이스로 제작된 음성 탐지 시스템 연구 저 자 : 정진호, 박우경, 이하영, 최수민, 홍준희, 김민수(지도교수) (중부대학교) 위 논문은 「2025 융합보안 ICT 컨퍼런스 및 하계학술 대회」에서 우수 논문으로 선정되어 이 상을 수여합니다. 2025년 7월 15일 한국전력공사 사 장 김 동 2025 한국융합보안학회 우수논문상	 제 2025-137호 상 장 우수상 팀명: 비상구 팀장: 최수민 팀원: 박우경, 이하영, 정진호, 홍준희 지도교수: 김민수 위 팀은 중부대학교 『2025-1학기 캡스톤디자인 경진대회』에서 우수한 성과[작품명: 인공지능을 이용한 딥보이스 탐지 프로그램]를 거두었기에 타의 모범이 되므로 이에 상장을 수여 합니다. 2025년 9월 10일 중부대학교 총장 이 정 열 2025-1 캡스톤디자인 경진대회 우수상
---	---

감사합니다.

QnA

