ML and DL approaches to identify CTC in IoMT communications

thesis coordinator
Francesco Benedetto

thesis advisor
Federica Massimi

candidate
Mario Cuomo, 569590

AGENDA

AGENDA

- CYBERATTACK TREND

AGENDA

- CYBERATTACK TREND
- OVERVIEW

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL
- GOAL TO ACHIEVE

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL
- GOAL TO ACHIEVE
- MACHINE LEARNING AND DEEP LEARNING ALGORITHMS

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL
- GOAL TO ACHIEVE
- MACHINE LEARNING AND DEEP LEARNING ALGORITHMS
- PROPOSED APPROACH

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL
- GOAL TO ACHIEVE
- MACHINE LEARNING AND DEEP LEARNING ALGORITHMS
- PROPOSED APPROACH
- RESULTS

AGENDA

- CYBERATTACK TREND
- OVERVIEW
- COVERT CHANNEL
  - COVERT TIMING CHANNEL
  - COVERT STORAGE CHANNEL
- GOAL TO ACHIEVE
- MACHINE LEARNING AND DEEP LEARNING ALGORITHMS
- PROPOSED APPROACH
- RESULTS
- CONCLUSION

CYBERATTACK TREND

CYBERATTACK TREND

© IBM report data breach 2022

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M
  - vulnerability in third-party software $ 4.55M

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M
  - vulnerability in third-party software $ 4.55M

- LIFECYCLE

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M
  - vulnerability in third-party software $ 4.55M

- LIFECYCLE
  - it took an average of 277 days to identify a breach!

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

compromised IoMT devices

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M
  - vulnerability in third-party software $ 4.55M

- LIFECYCLE
  - it took an average of 277 days to identify a breach!

CYBERATTACK TREND

© IBM report data breach 2022

- HEALTHCARE GETS HIT HARD
  - up 42% since 2020
  - $ 10.10M

compromised IoMT devices

- STOLEN OR COMPROMISED CREDENTIALS
  - phishing $ 4.91M
  - business email compromised $ 4.89M
  - vulnerability in third-party software $ 4.55M

- LIFECYCLE
  - it took an average of 277 days to identify a breach!

CTC channel attack

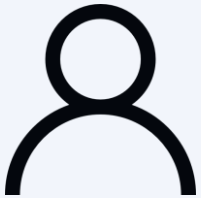
OVERVIEW

OVERVIEW

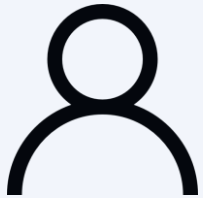
ALICE

BOB

OVERVIEW

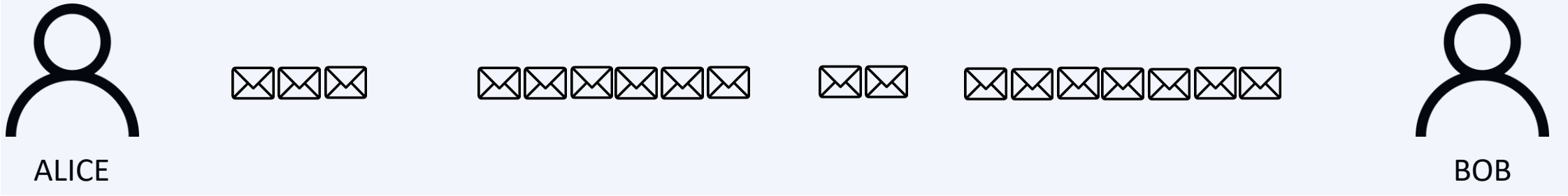


ALICE

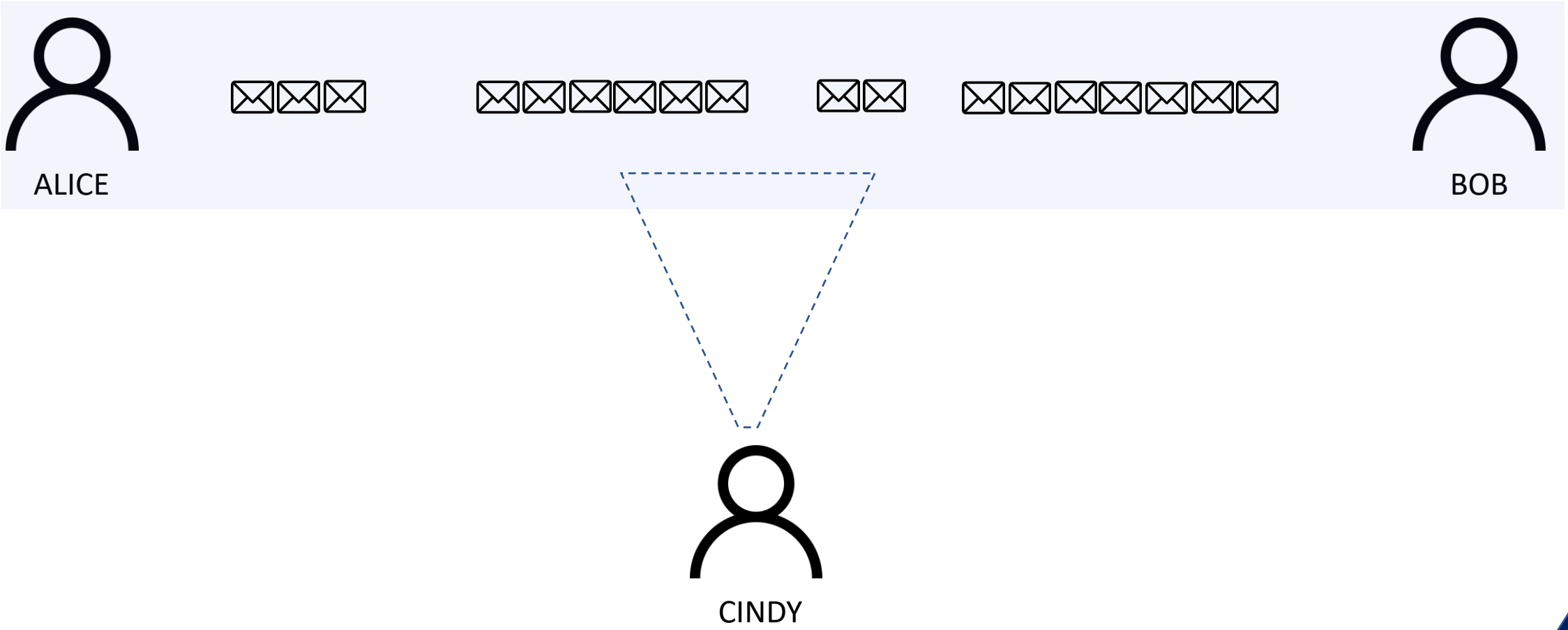


BOB

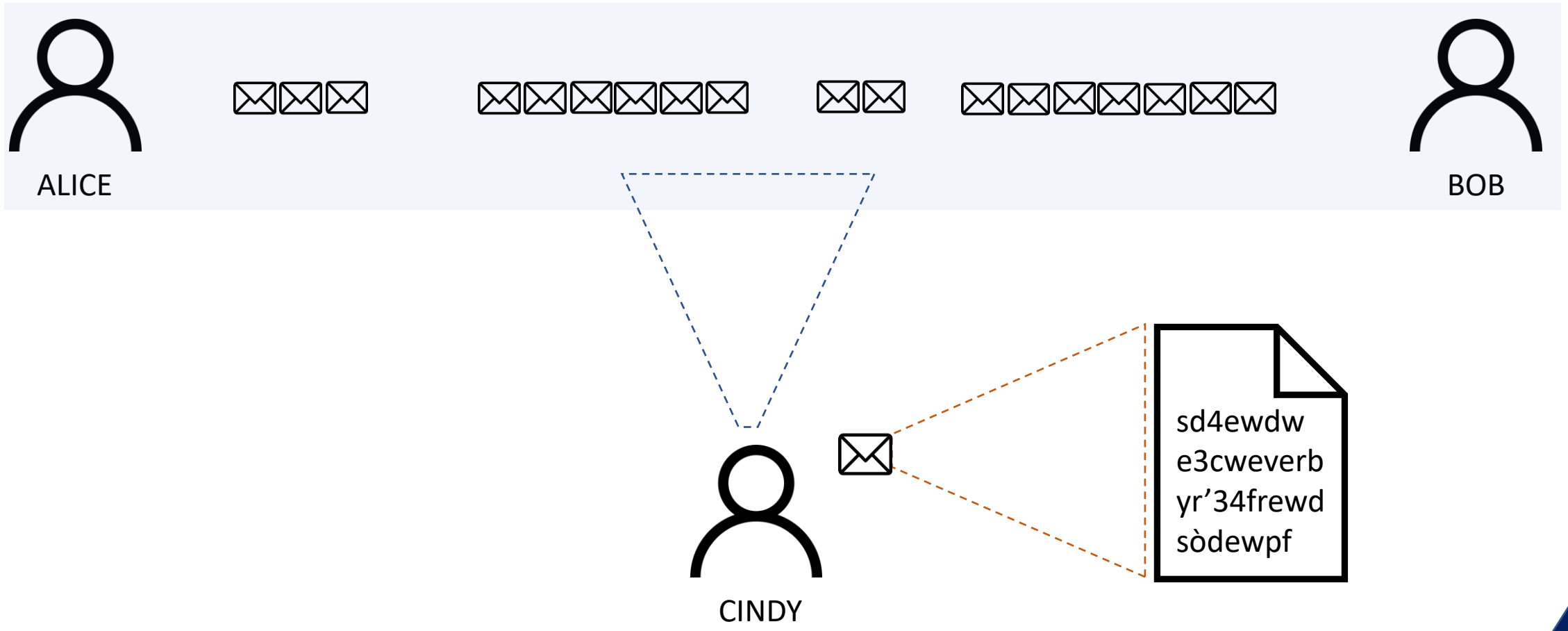
OVERVIEW



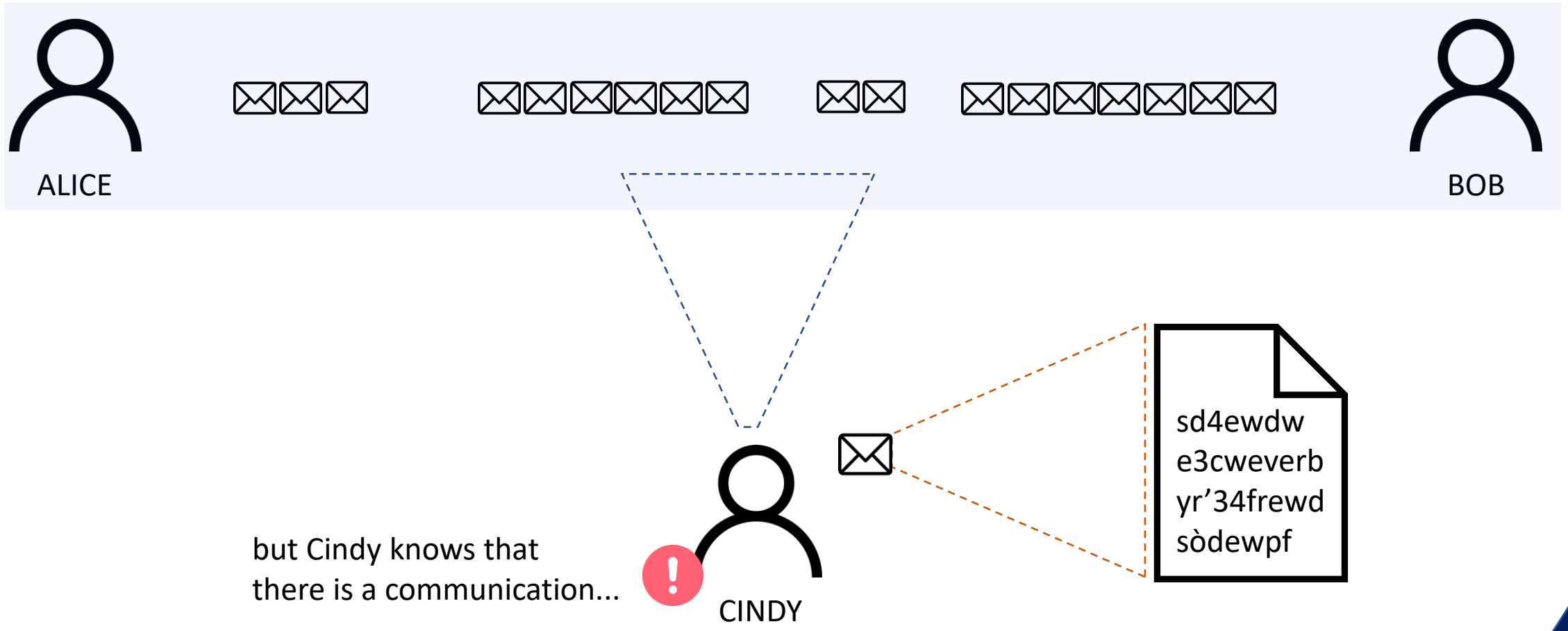
OVERVIEW



OVERVIEW



OVERVIEW



COVERT CHANNEL

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT STORAGE CHANNEL

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT STORAGE CHANNEL

communicate by modifying a "storage location", such as a hard drive

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT STORAGE CHANNEL

communicate by modifying a "storage location", such as a hard drive

COVERT TIMING CHANNEL

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT STORAGE CHANNEL

communicate by modifying a "storage location", such as a hard drive

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver

COVERT CHANNEL

“channels not intended for information transfer at all, such as the service program's effect on system load”

B. Lampson, 1973

COVERT STORAGE CHANNEL

communicate by modifying a "storage location", such as a hard drive

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver

CTC TIME-REPLAY - *Serdar Cabuk «Network covert channels: Design, analysis, detection, and elimination»*

COVERT TIMING CHANNEL

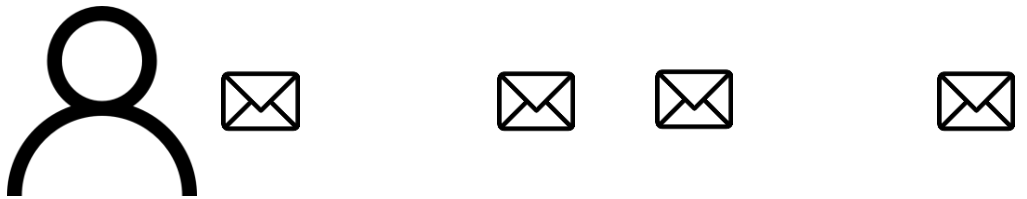
perform operations that affect the "real response time observed" by the receiver

ALICE

BOB

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver



COVERT TIMING CHANNEL

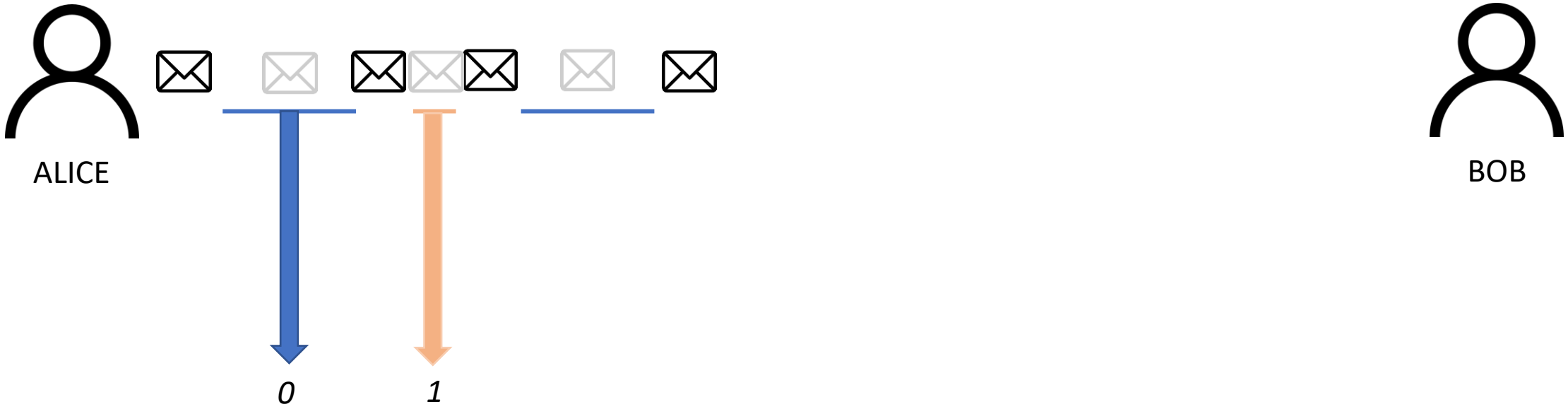
perform operations that affect the "real response time observed" by the receiver

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver

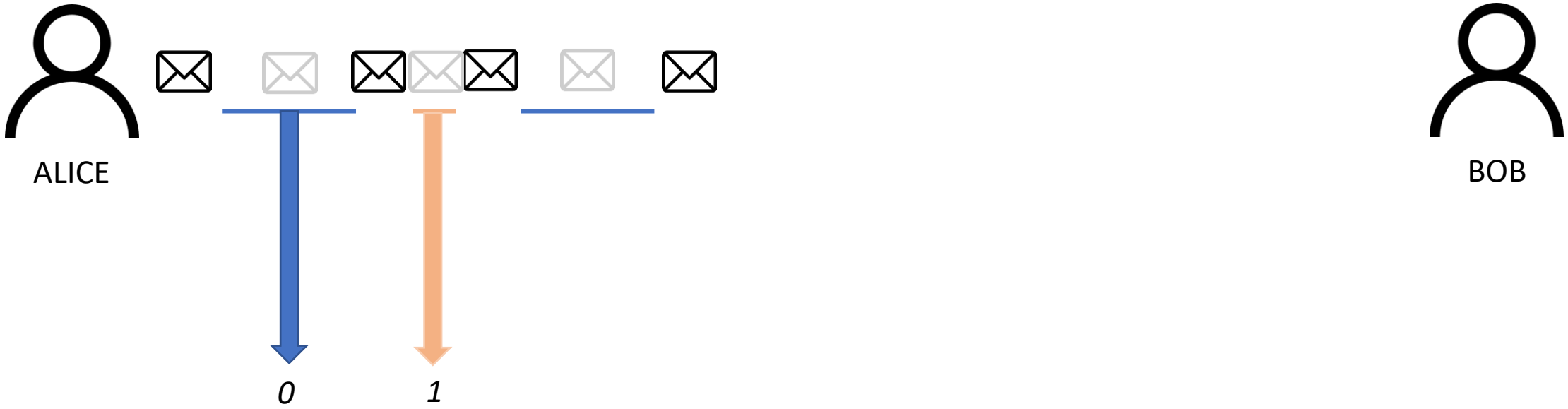
COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver



COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver

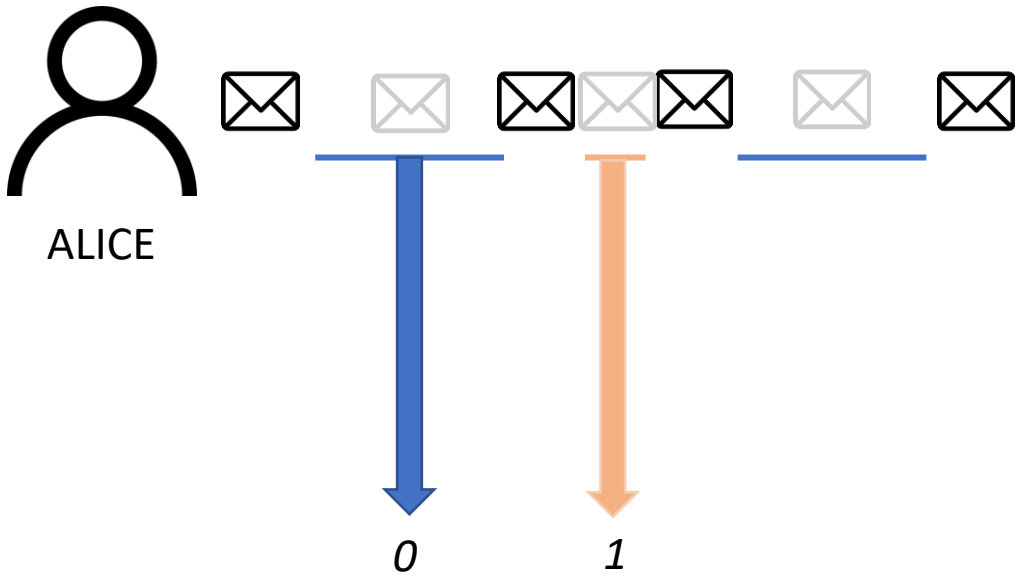


covering message

covert message

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver

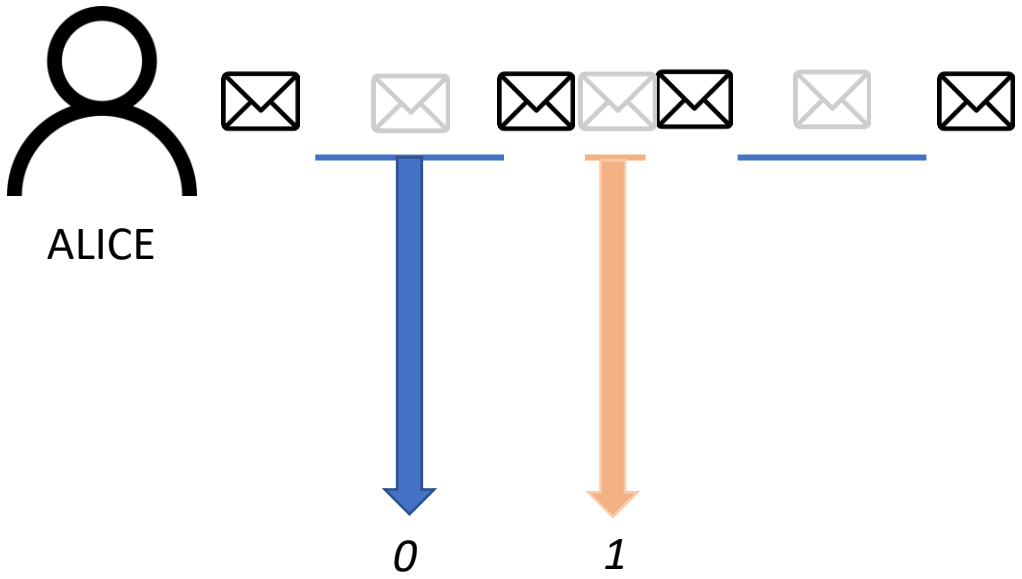


covering message

covert message

COVERT TIMING CHANNEL

perform operations that affect the "real response time observed" by the receiver



Cindy knows covering message
but not covert message!

covering message

covert message

GOAL TO ACHIEVE

GOAL TO ACHIEVE

ALICE

BOB

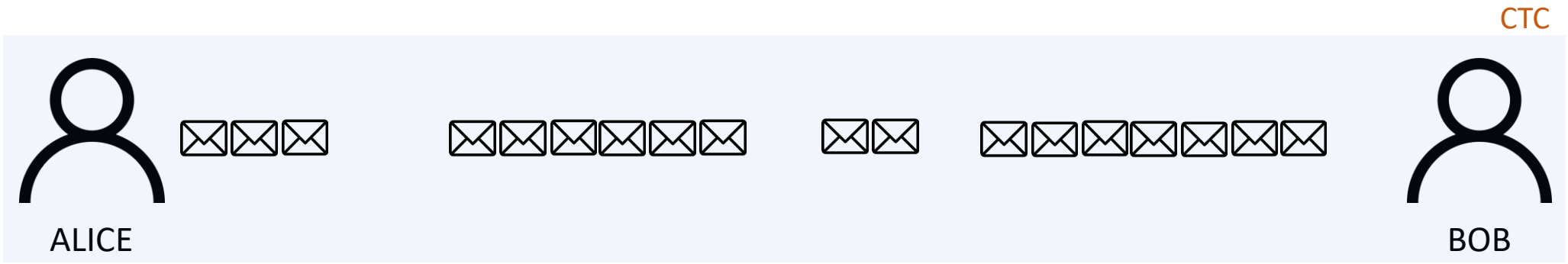
GOAL TO ACHIEVE

CTC

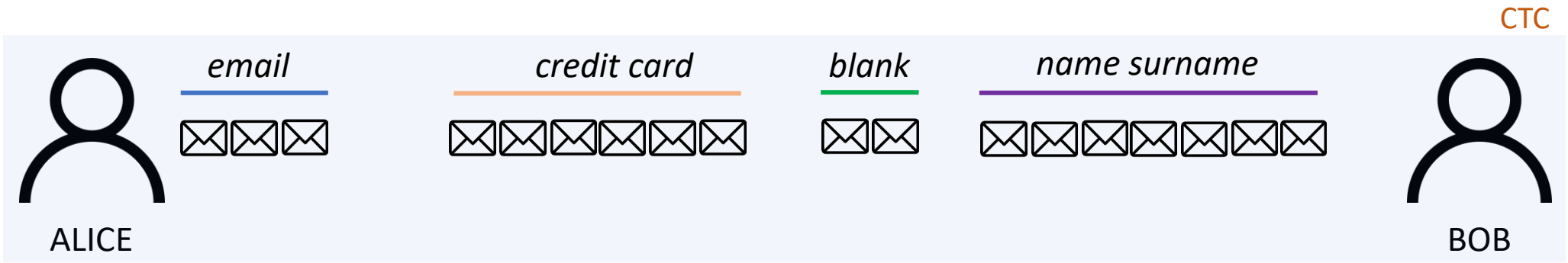
ALICE

BOB

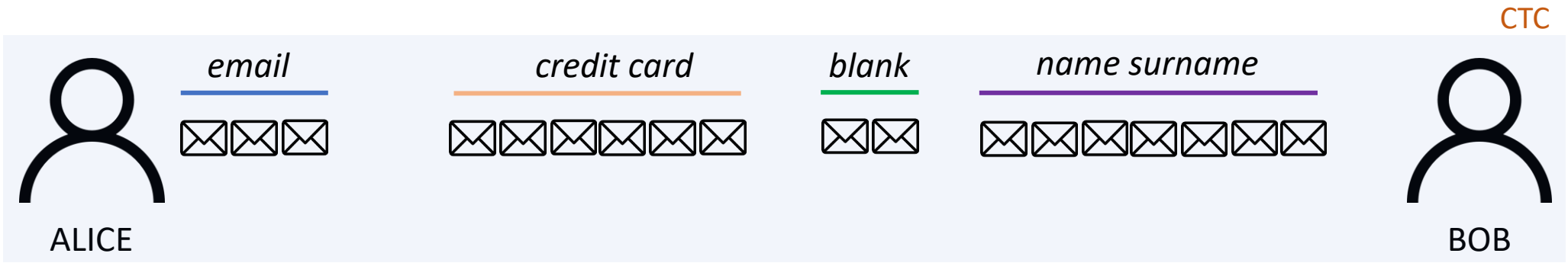
GOAL TO ACHIEVE

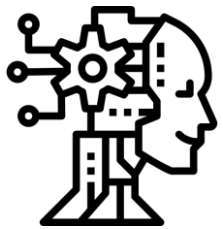


GOAL TO ACHIEVE

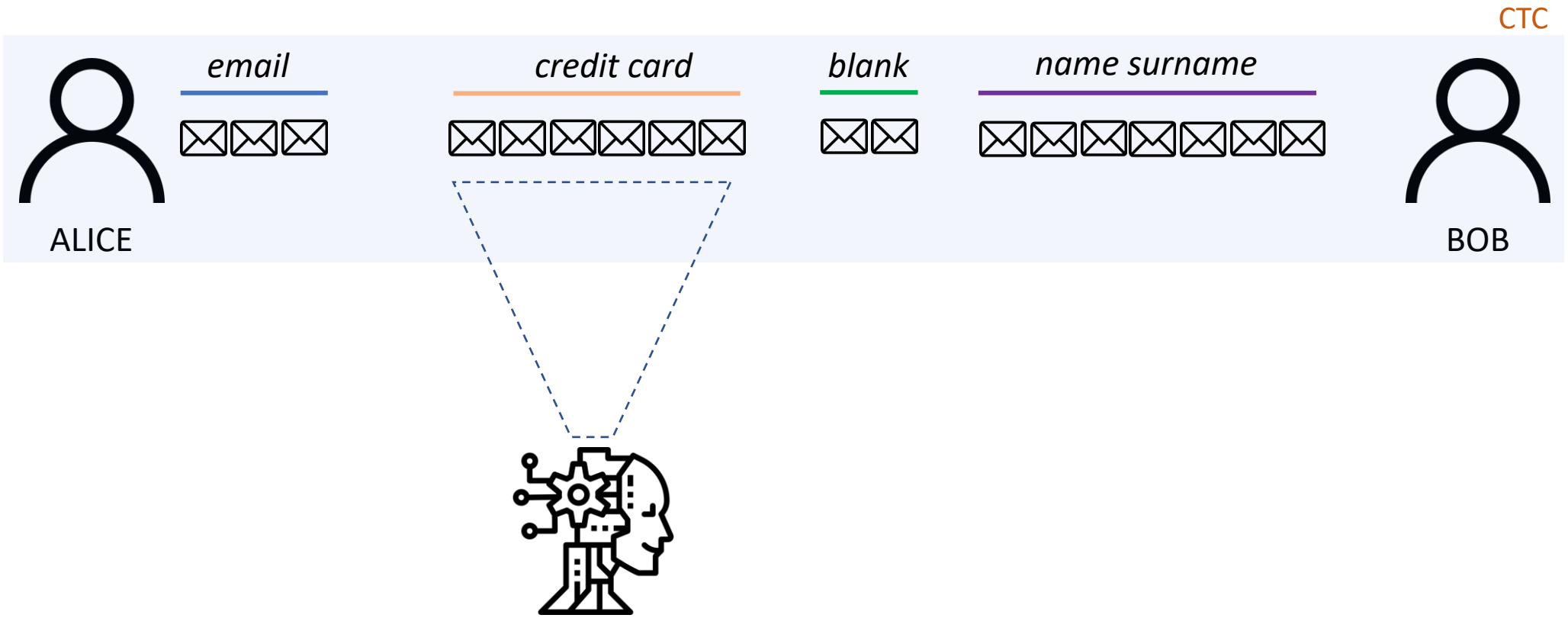


GOAL TO ACHIEVE

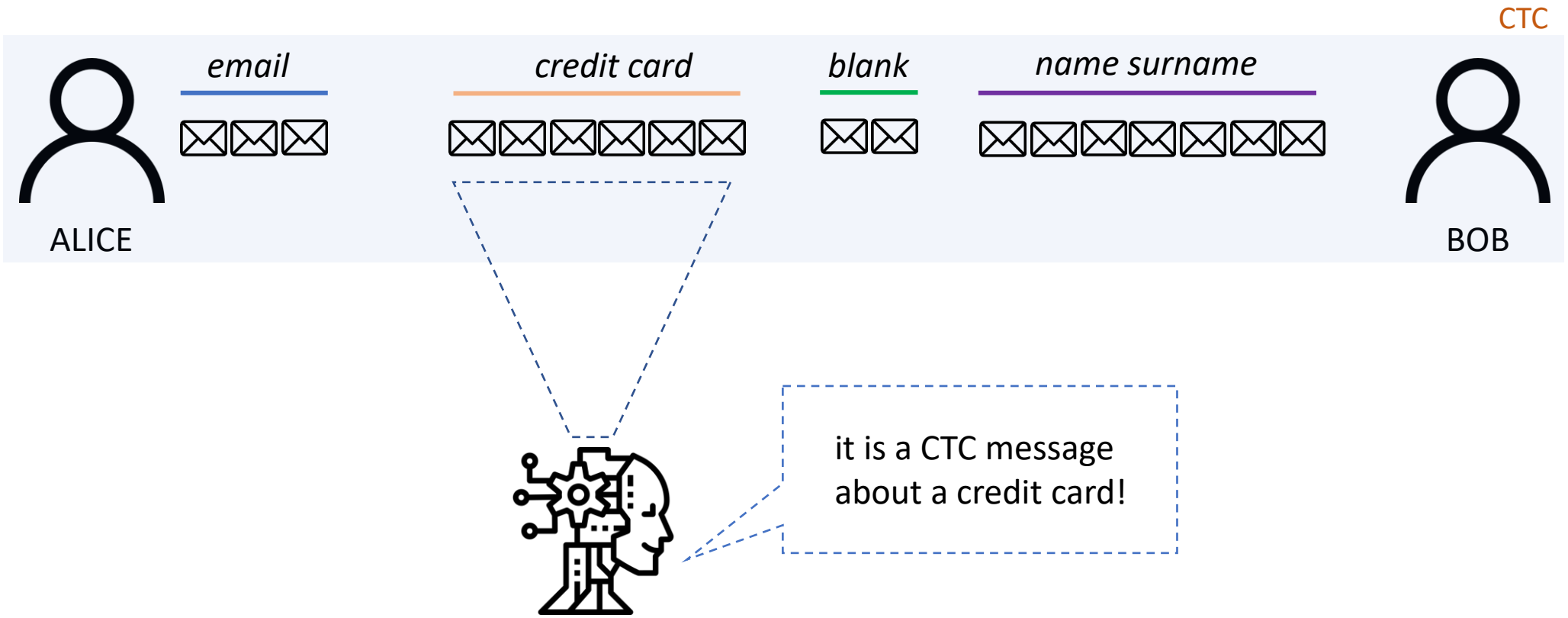




GOAL TO ACHIEVE



GOAL TO ACHIEVE



PROPOSED APPROACH

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

- CONVOLUTIONAL NEURAL NETWORK

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

- CONVOLUTIONAL NEURAL NETWORK
- SIAMESE NEURAL NETWORK

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

- CONVOLUTIONAL NEURAL NETWORK
- SIAMESE NEURAL NETWORK
- K-MEANS

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

- CONVOLUTIONAL NEURAL NETWORK
- SIAMESE NEURAL NETWORK
- K-MEANS
- RANDOM FOREST

PROPOSED APPROACH

- LAB KATHARÁ - *open source container-based network emulation system* (Roma Tre)
- CTC IMPLEMENTATION (python)
- 500 instances for each class
  - 400 as training set
  - 100 as validation set

- CONVOLUTIONAL NEURAL NETWORK
- SIAMESE NEURAL NETWORK
- K-MEANS
- RANDOM FOREST

python – sklearn and tensorflow, keras

THE BIG PICTURE

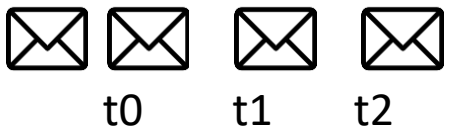
STEP ONE

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS



STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

STEP ONE – PACKETS TO SPECTROGRAMS

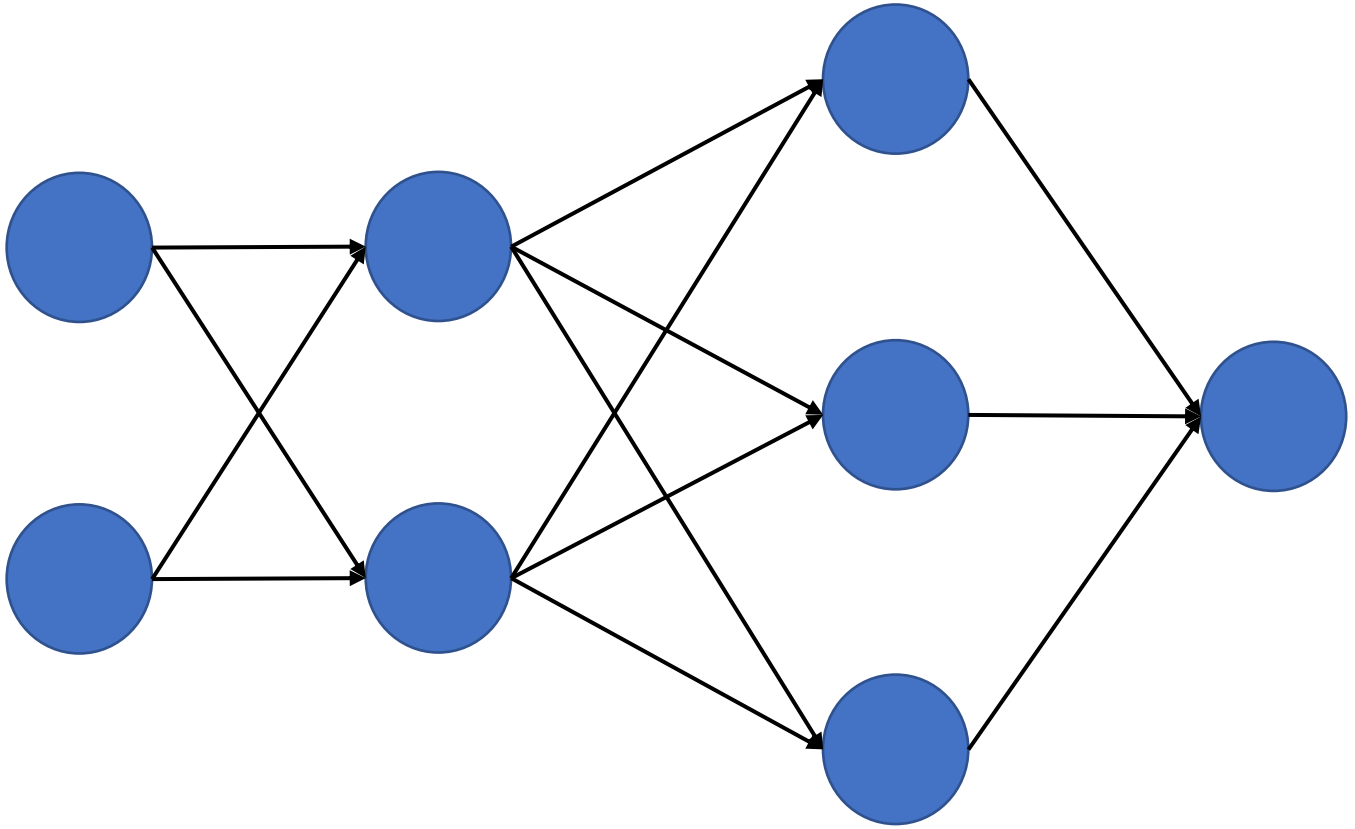
S. Al-Eidi, O. Darwish, Y. Chen and G. Husari, "SnapCatch: Automatic Detection of Covert Timing Channels Using Image Processing and Machine Learning"

THE BIG PICTURE

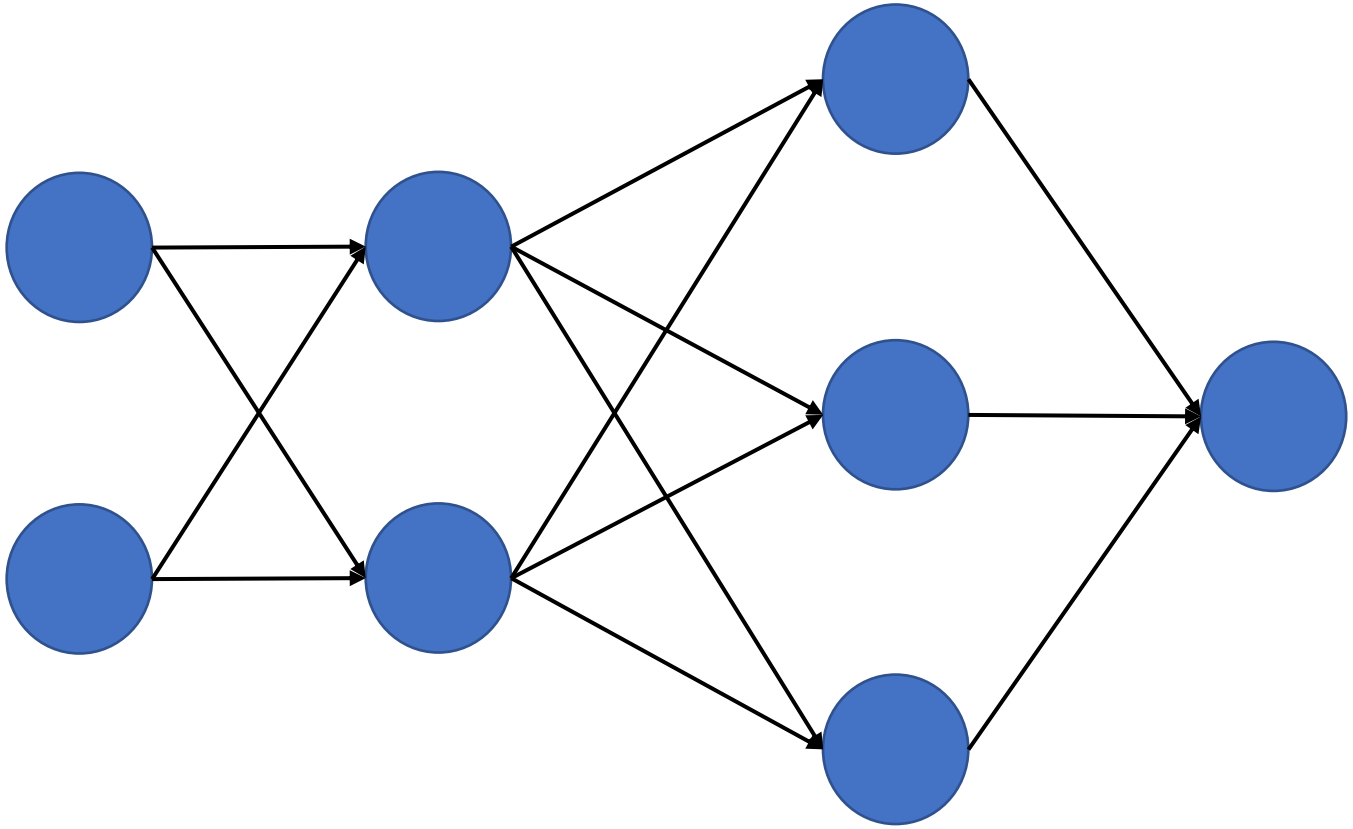
STEP TWO

STEP TWO – CONVOLUTION NEURAL NETWORK

STEP TWO – CONVOLUTION NEURAL NETWORK

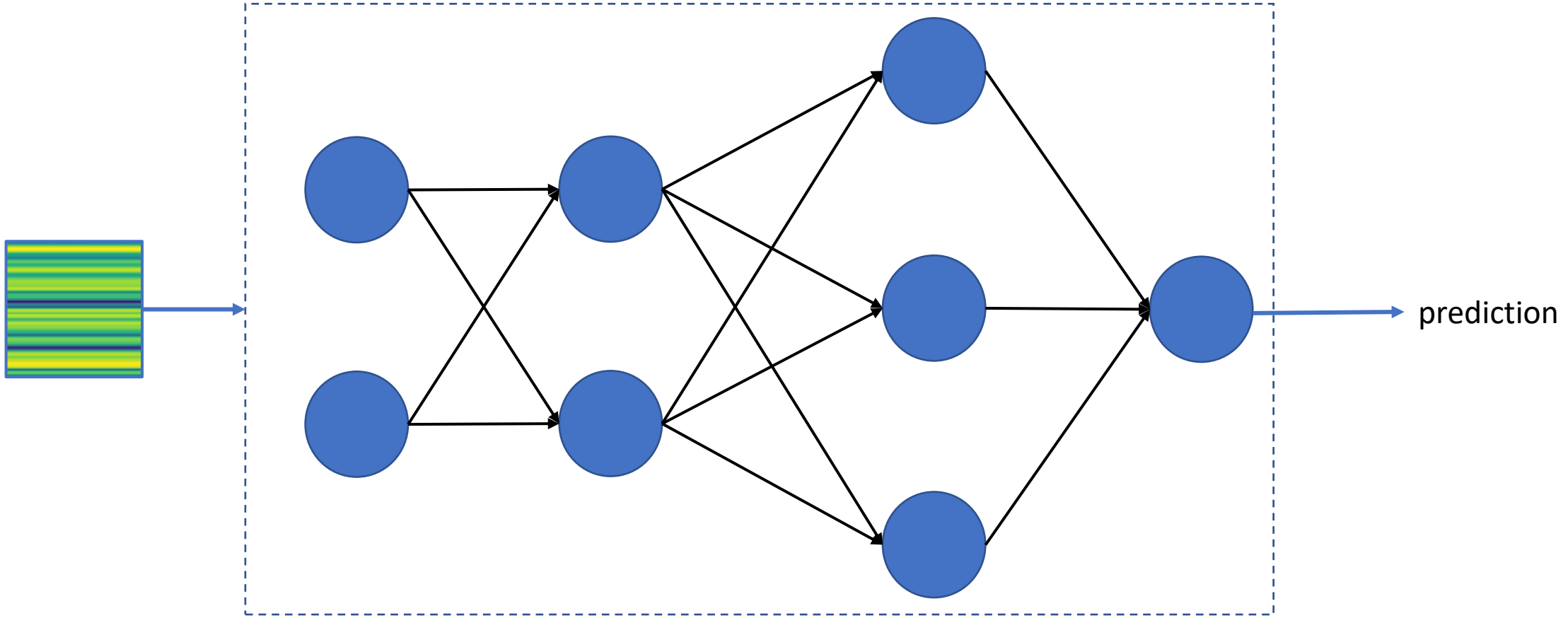


STEP TWO – CONVOLUTION NEURAL NETWORK



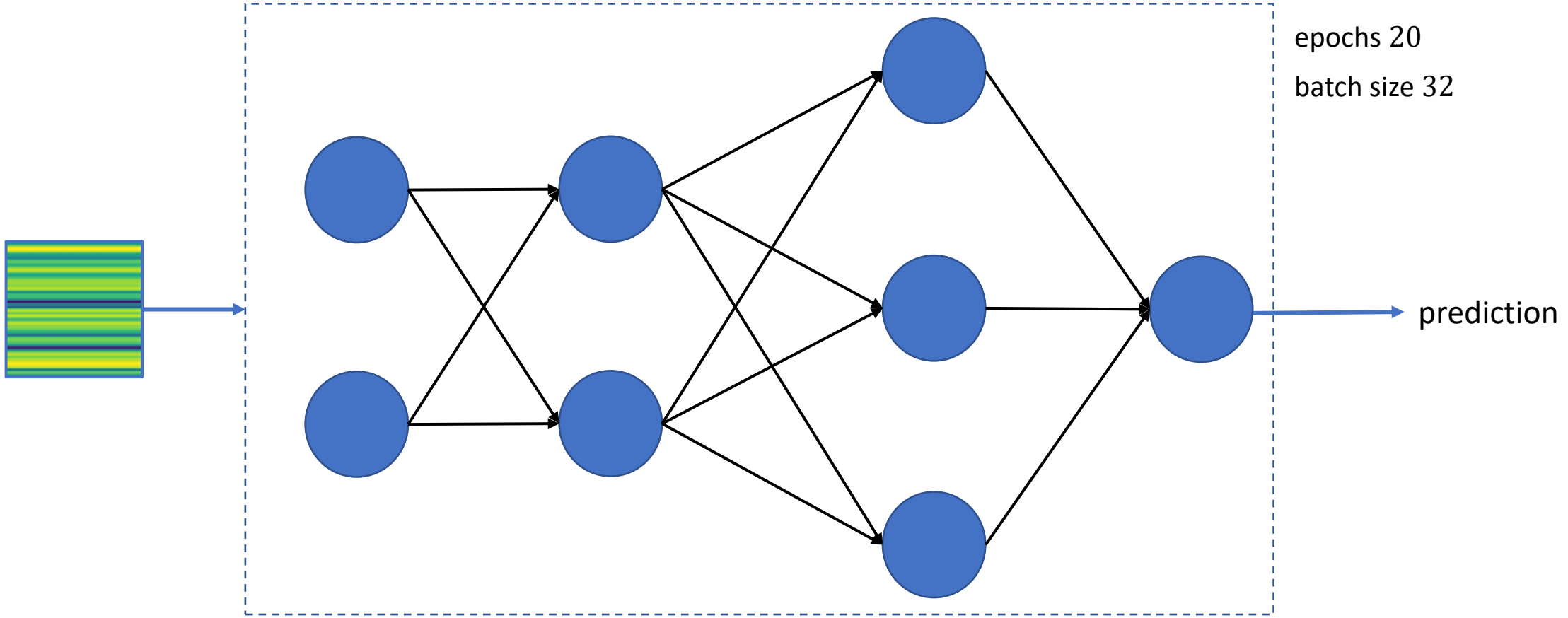
input size $224 \times 224 \times 3$	max pooling 2×2	fully connected 128 units
features map 3×3	dropout $20\ \%$	output softmax 4 units

STEP TWO – CONVOLUTION NEURAL NETWORK



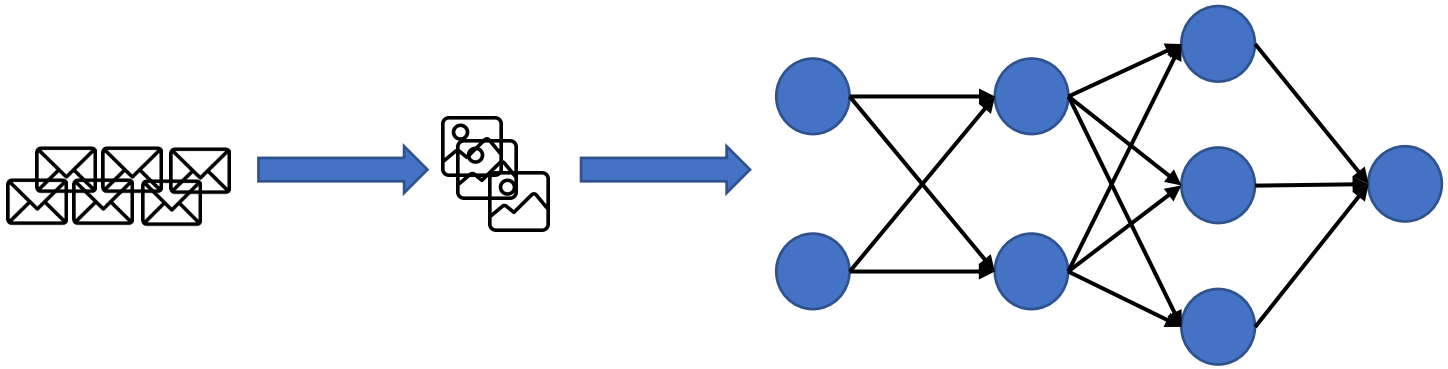
input size $224 \times 224 \times 3$	max pooling 2×2	fully connected 128 units
features map 3×3	dropout $20\ \%$	output softmax 4 units

STEP TWO – CONVOLUTION NEURAL NETWORK



input size $224 \times 224 \times 3$	max pooling 2×2	fully connected 128 units
features map 3×3	dropout 20 %	output softmax 4 units

THE BIG PICTURE



CONFUSION MATRIX CNN

	cc	*email*	*ns*	*blank*
cc	65	9	26	0
email	0	100	0	0
ns	16	5	73	6
blank	0	0	1	99

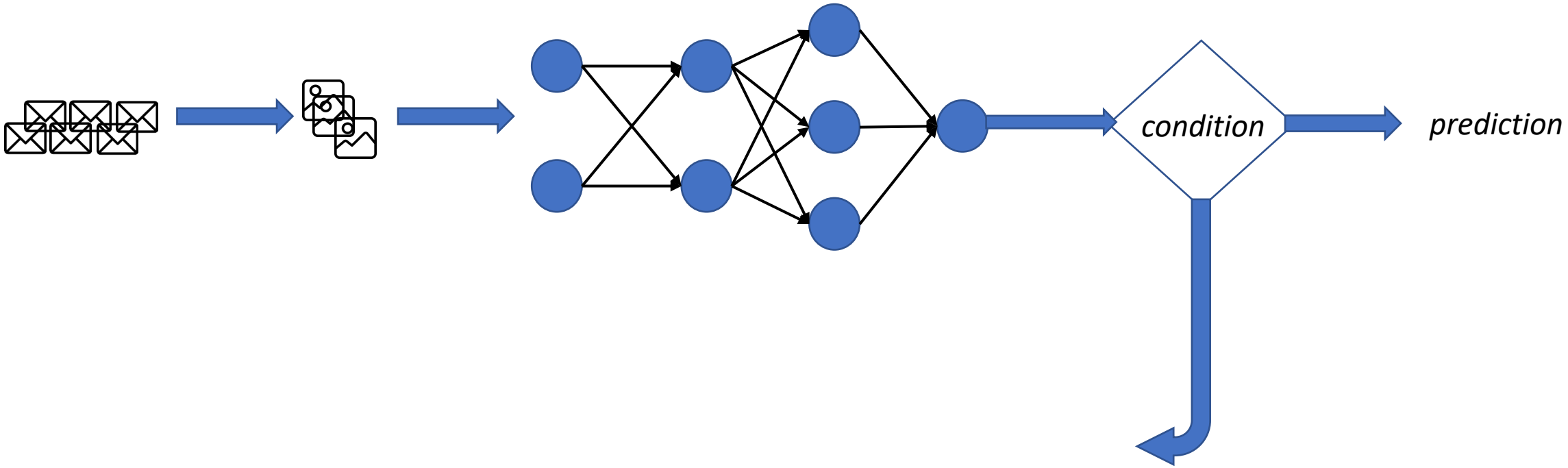
cc – credit card
ns – name surname

CONFUSION MATRIX CNN

	cc	*email*	*ns*	*blank*
cc	65	9	26	0
email	0	100	0	0
ns	16	5	73	6
blank	0	0	1	99

cc – credit card
ns – name surname

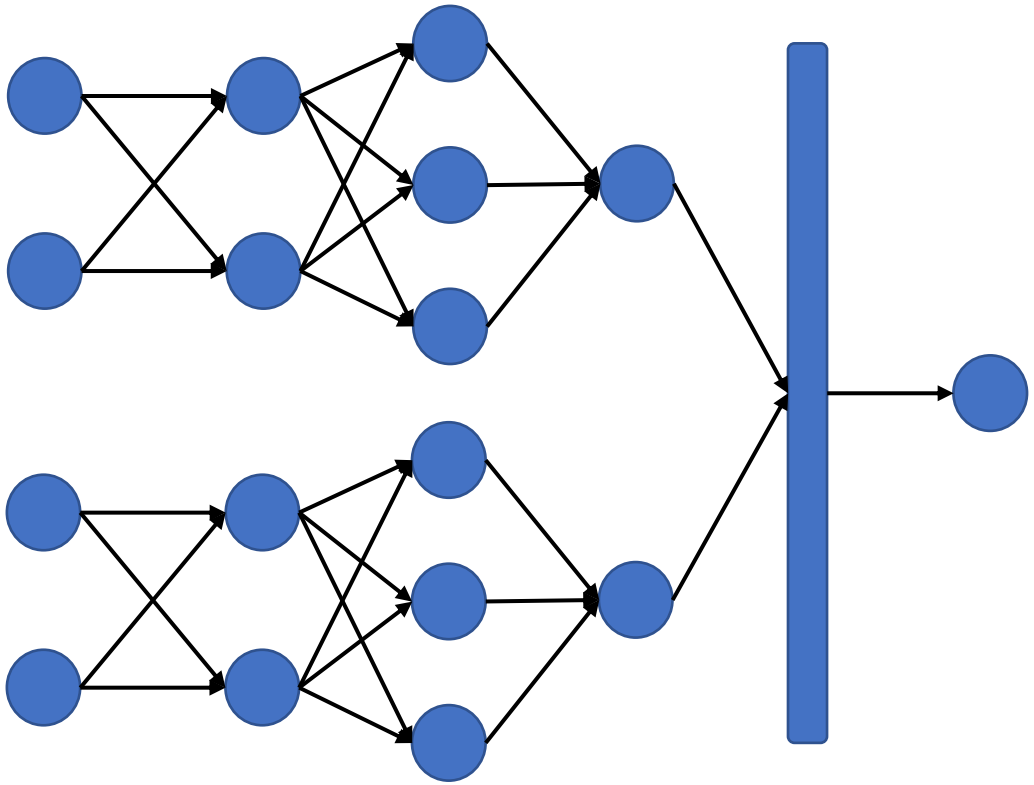
THE BIG PICTURE



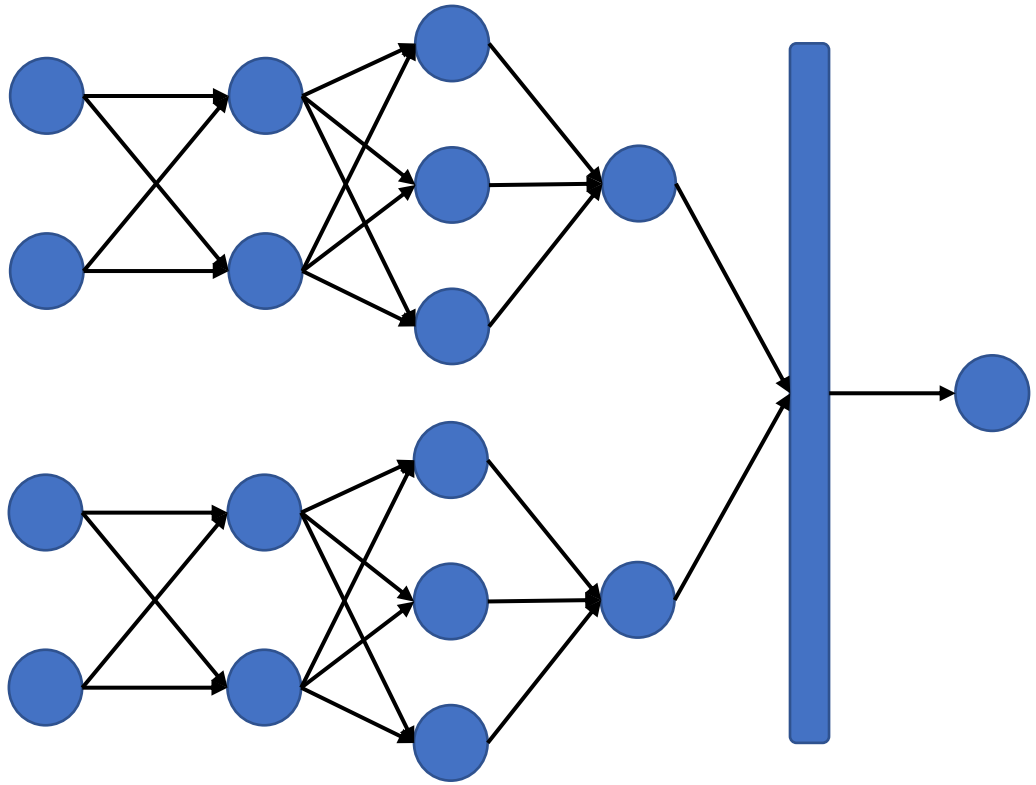
STEP THREE

STEP THREE – SIAMESE NEURAL NETWORK

STEP THREE – SIAMESE NEURAL NETWORK



STEP THREE – SIAMESE NEURAL NETWORK



input size $2 \times 224 \times 224 \times 3$

reflection pad 2D

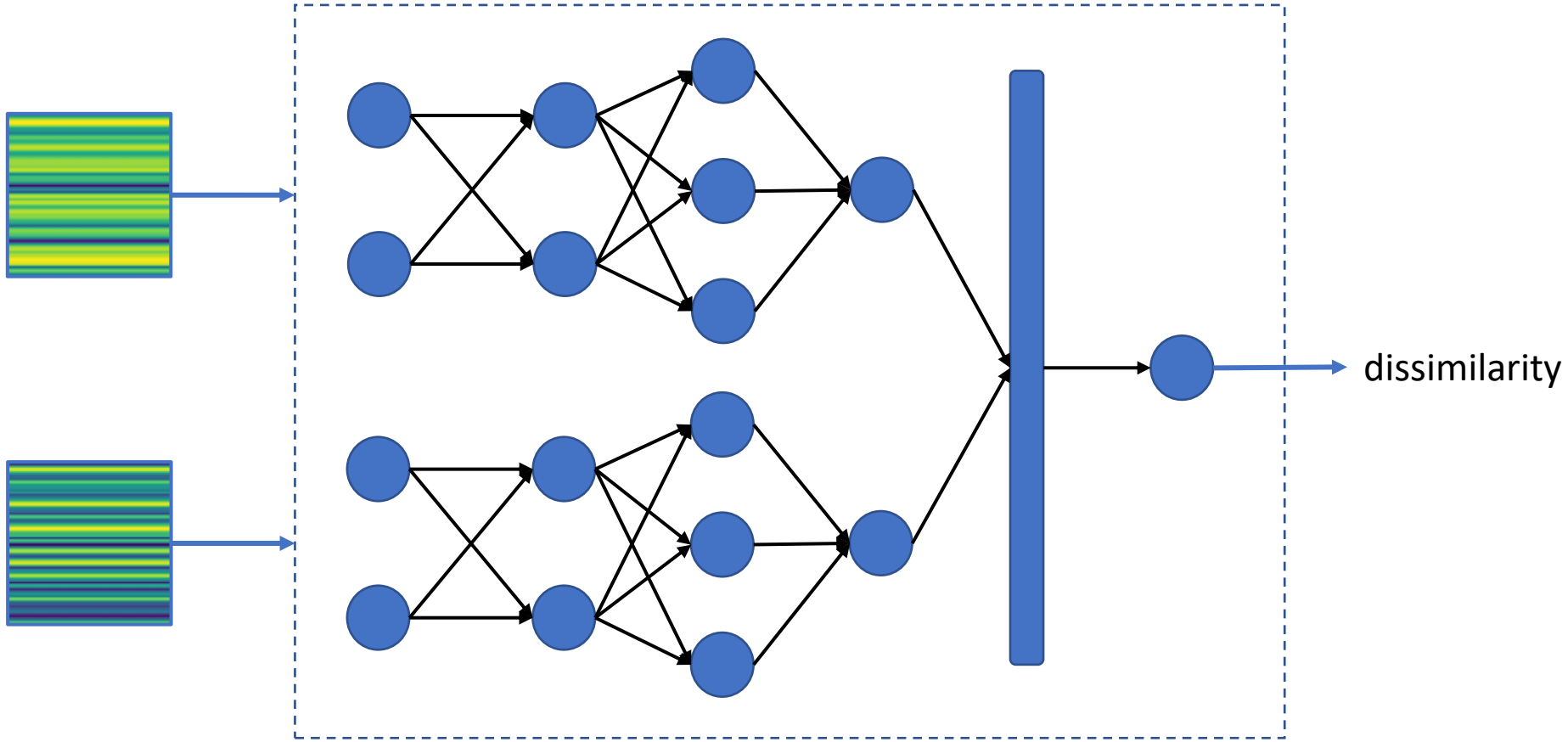
3 convolutional layers

relu as activation

batch normalization

sigmoid 1 unit

STEP THREE – SIAMESE NEURAL NETWORK



input size $2 \times 224 \times 224 \times 3$

reflection pad 2D

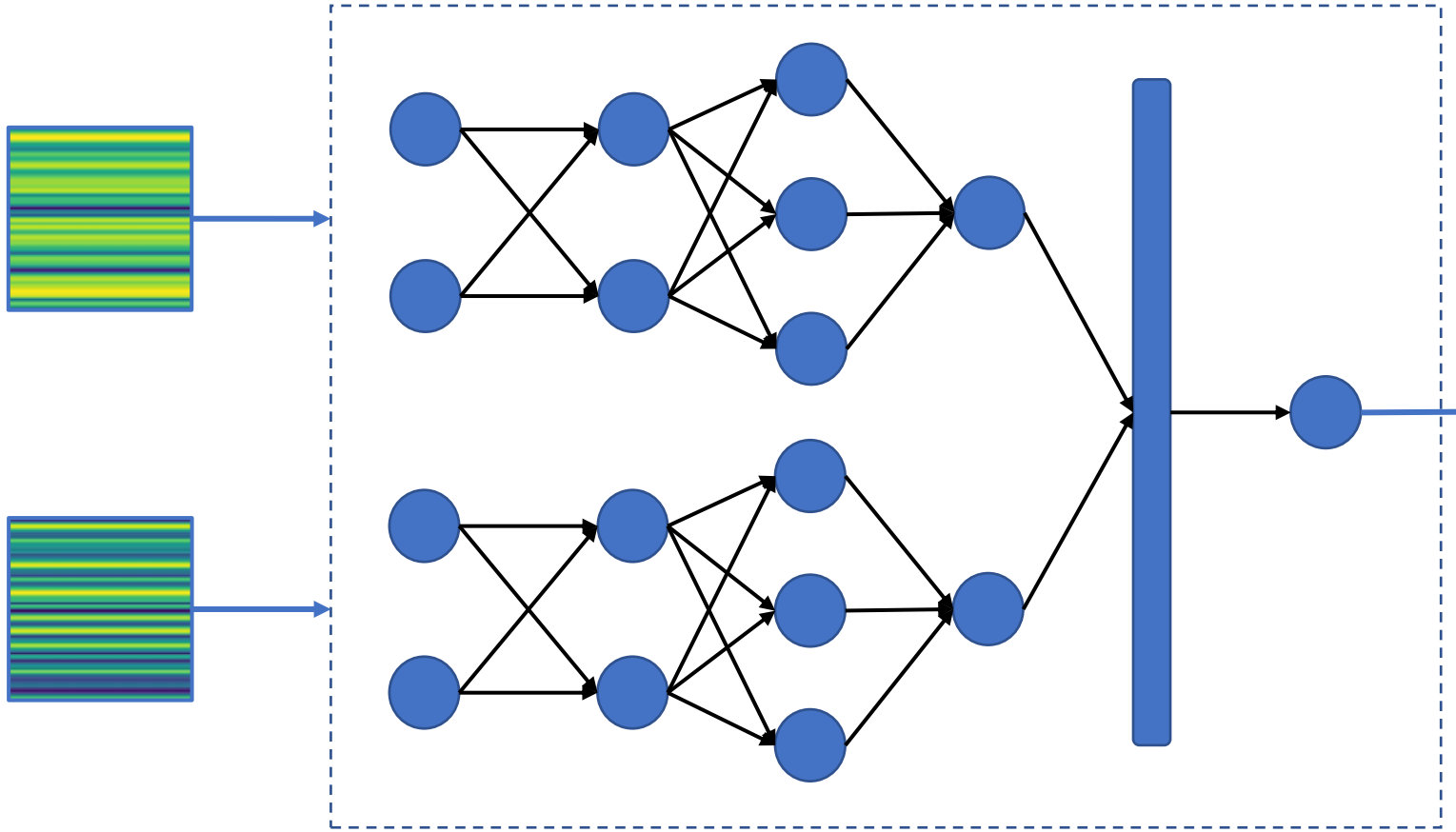
3 convolutional layers

relu as activation

batch normalization

sigmoid 1 unit

STEP THREE – SIAMESE NEURAL NETWORK



epochs 100

batch size 64

dissimilarity

input size $2 \times 224 \times 224 \times 3$

reflection pad 2D

3 convolutional layers

relu as activation

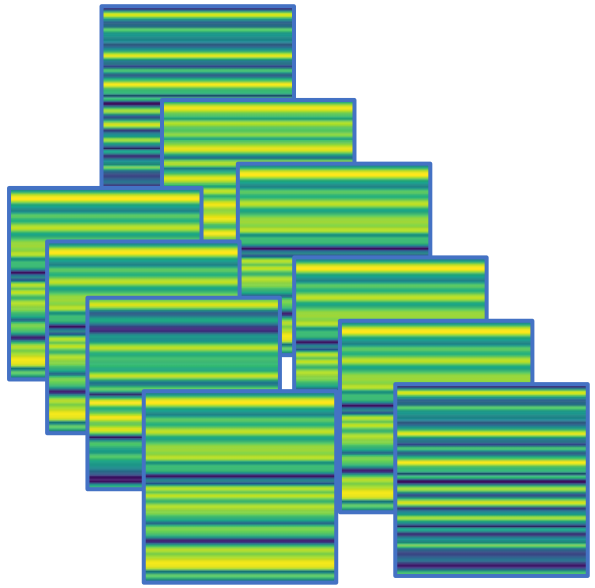
batch normalization

sigmoid 1 unit

STEP FOUR

STEP FOUR – PROTOTYPE SELECTION

STEP FOUR – PROTOTYPE SELECTION



STEP FOUR – PROTOTYPE SELECTION

prototype selection

using k-means

STEP FOUR – PROTOTYPE SELECTION

prototype selection

using k-means

prototype of a credit card spectrogram

prototype of a name surname spectrogram

STEP FOUR – PROTOTYPE SELECTION

prototype selection

using k-means

prototype of a credit card spectrogram

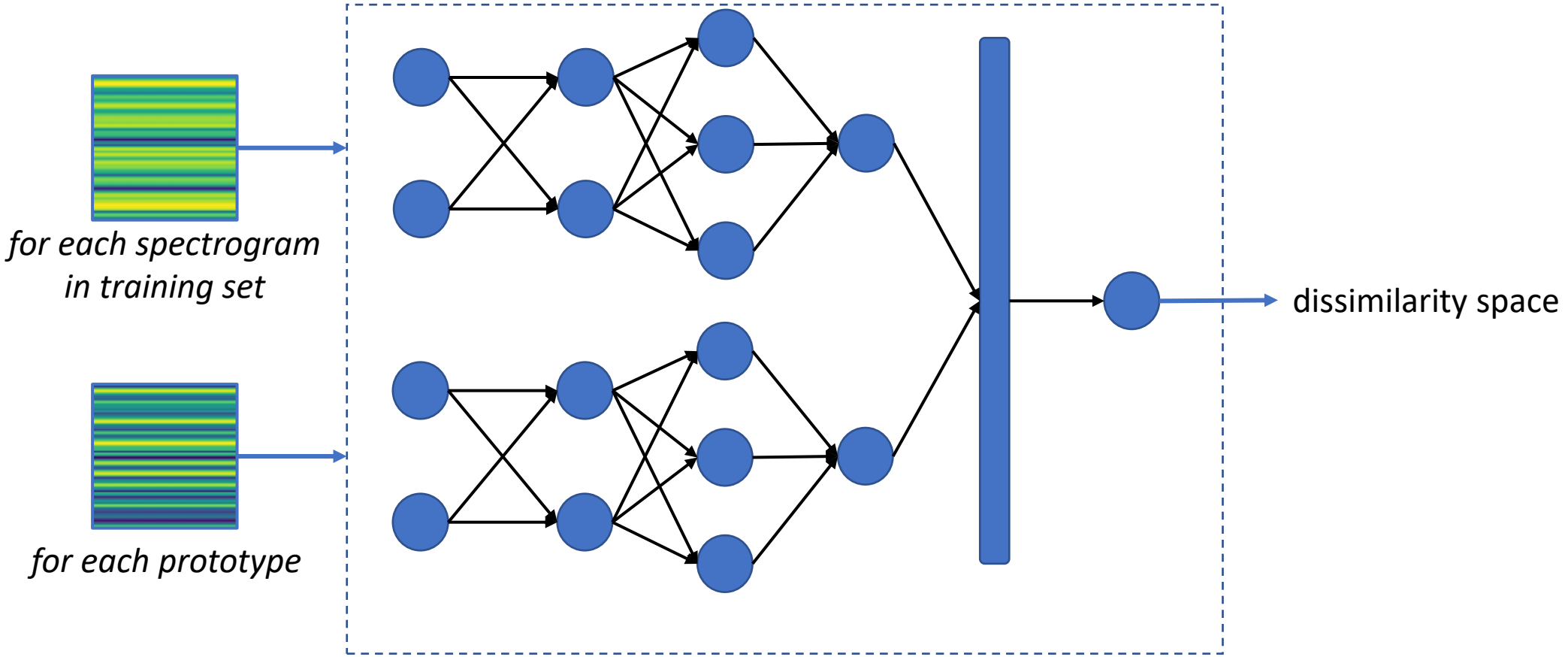
prototype of a name surname spectrogram

Loris Nanni et al. «Spectrogram Classification Using Dissimilarity Space».

STEP FIVE

STEP FIVE – DISSIMILARITY SPACE DATASET

STEP FIVE – DISSIMILARITY SPACE DATASET

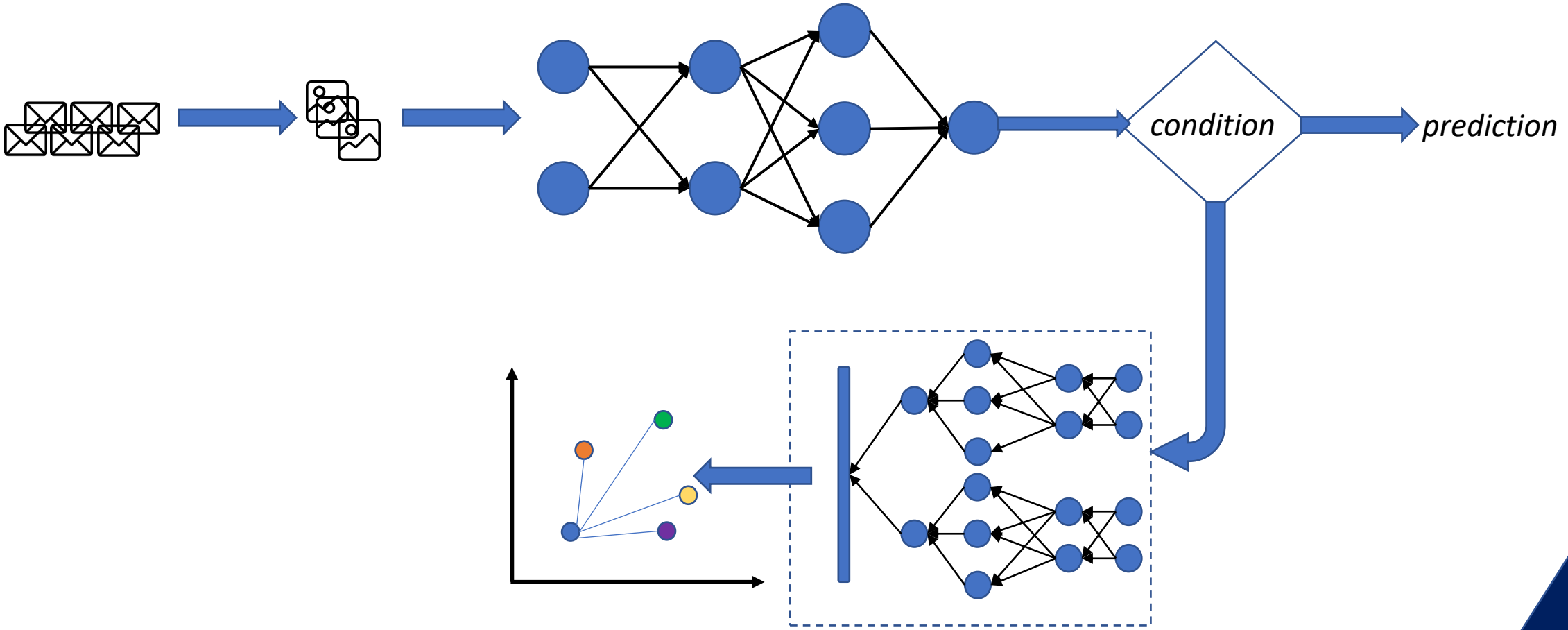


STEP FIVE – DISSIMILARITY SPACE DATASET

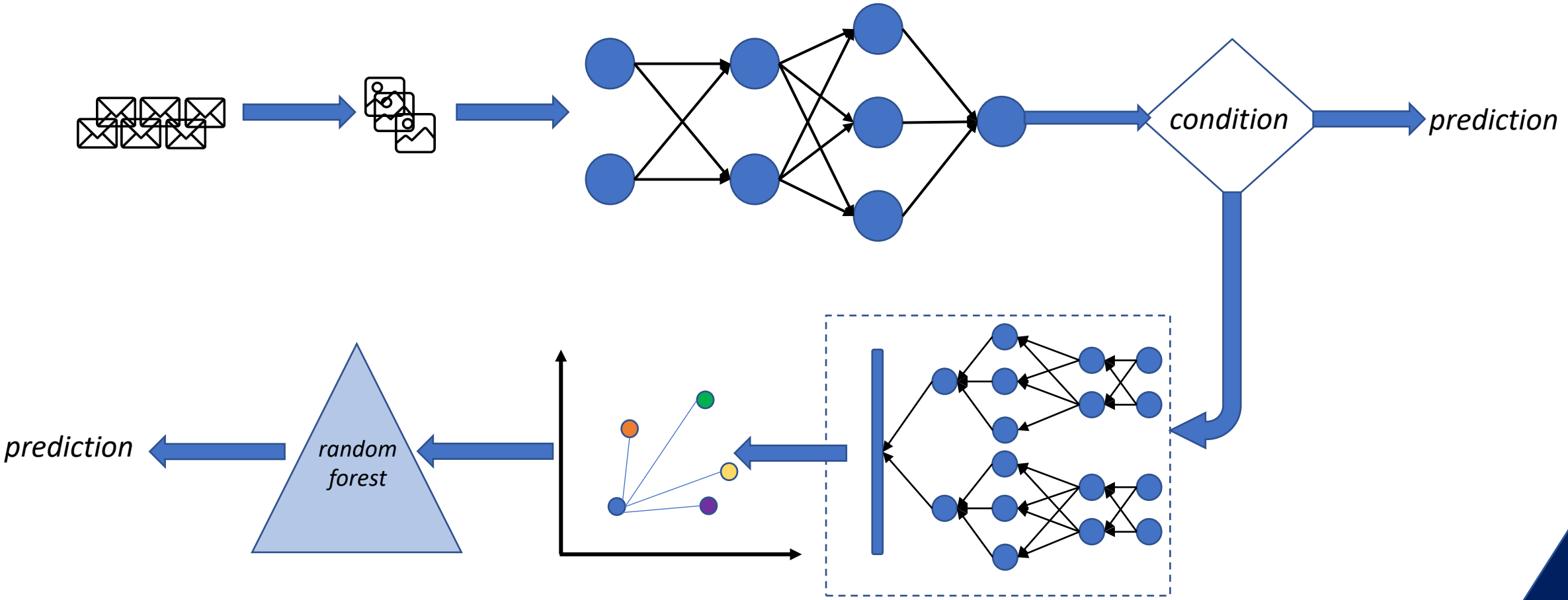
distance to cc	*distance to ns*	*class*
0.22	0.67	*credit card*
0.1	0.97	*name surname*
0.54	0.3	*credit card*
0.44	0.67	*credit card*
…	…	…

cc – credit card
ns – name surname

THE BIG PICTURE

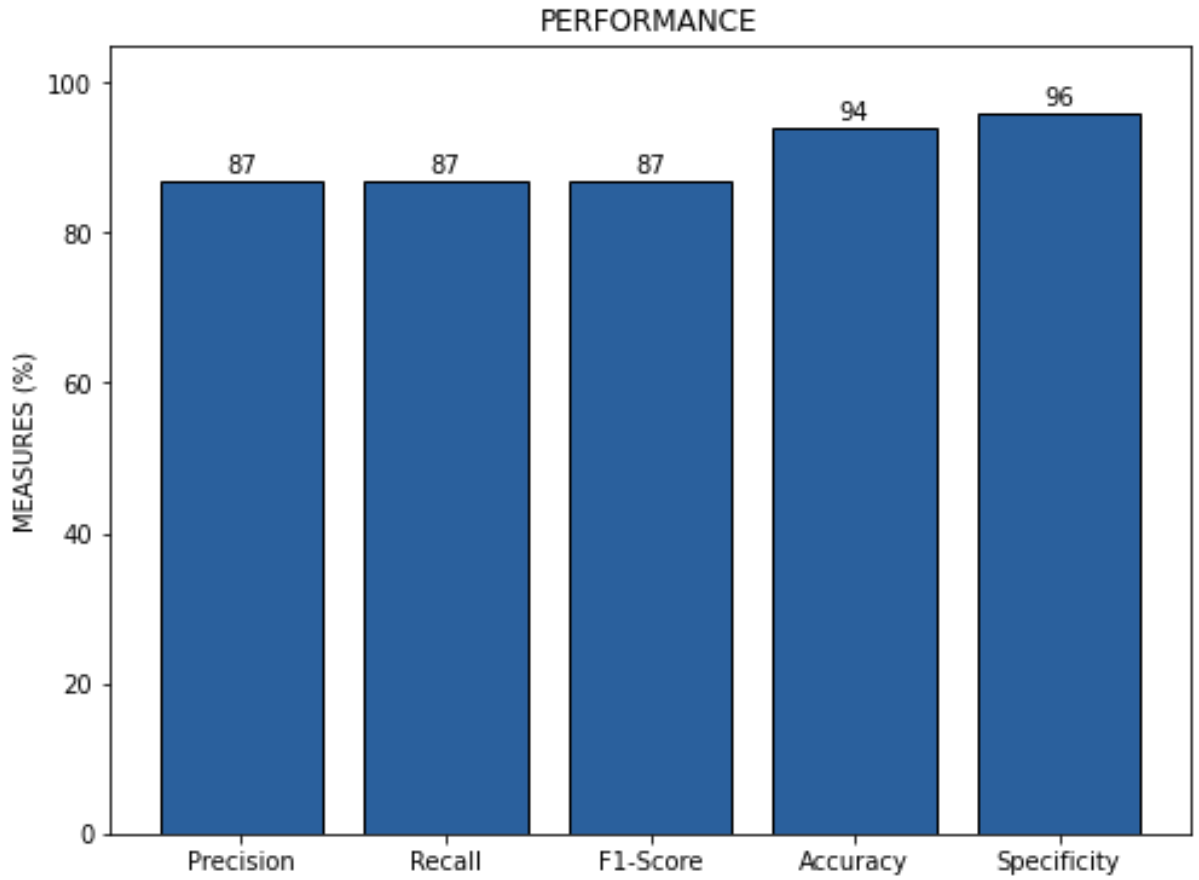


THE BIG PICTURE

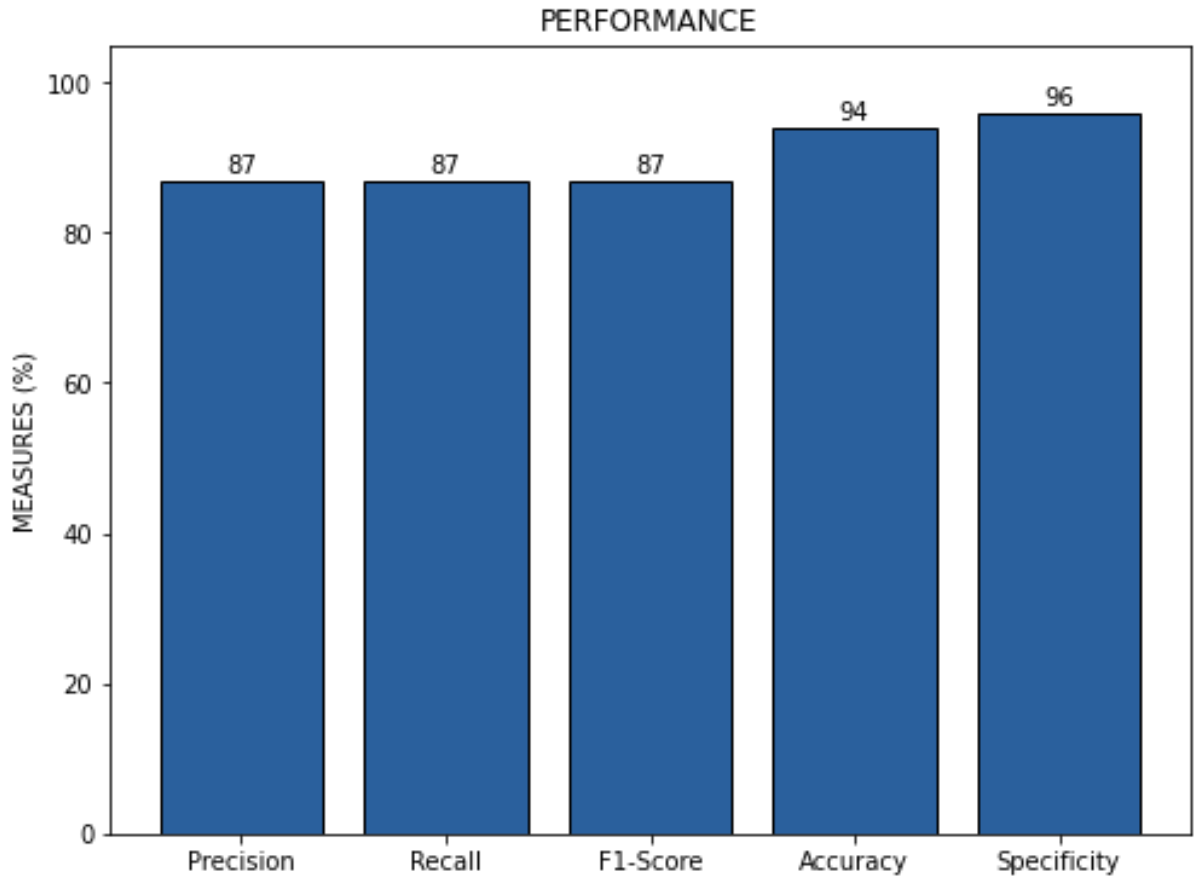


RESULTS

RESULTS

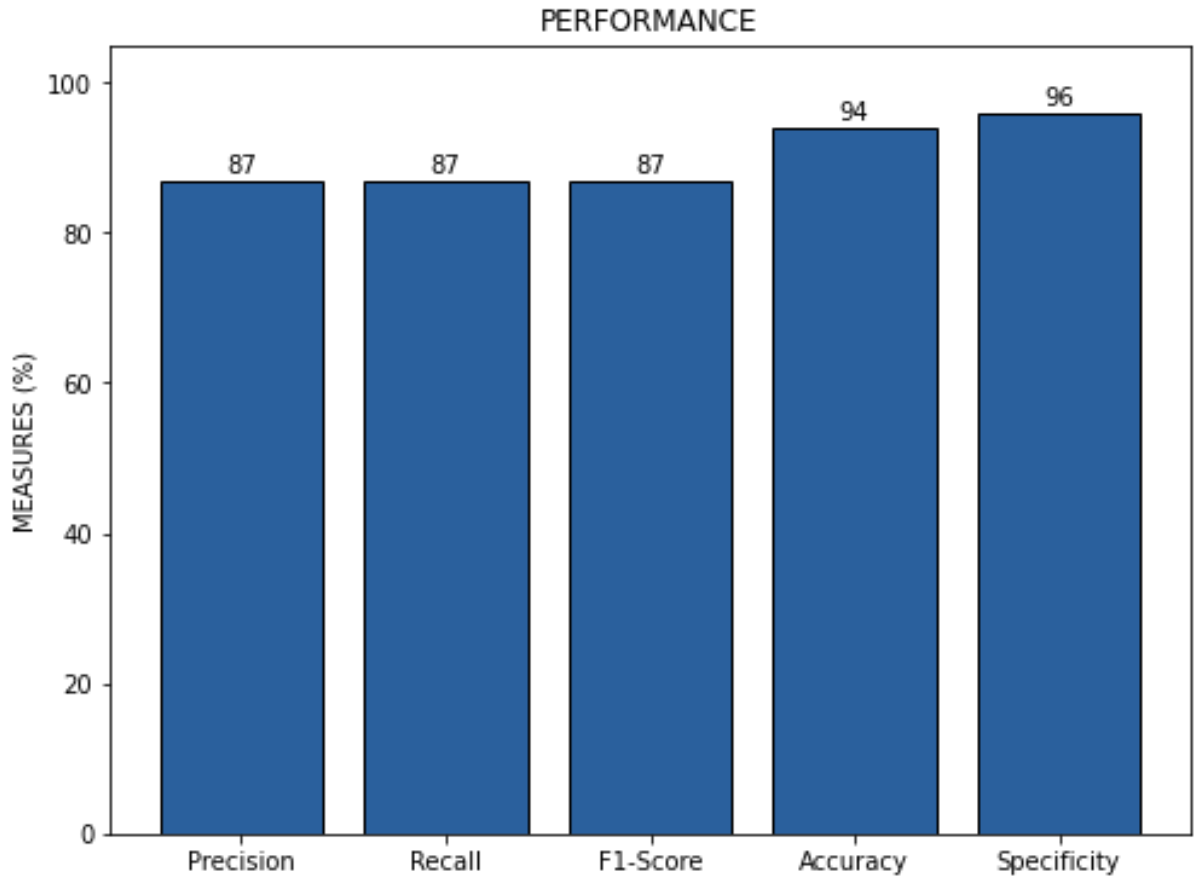


RESULTS



SOME STATE OF ART ACCURACY RESULTS

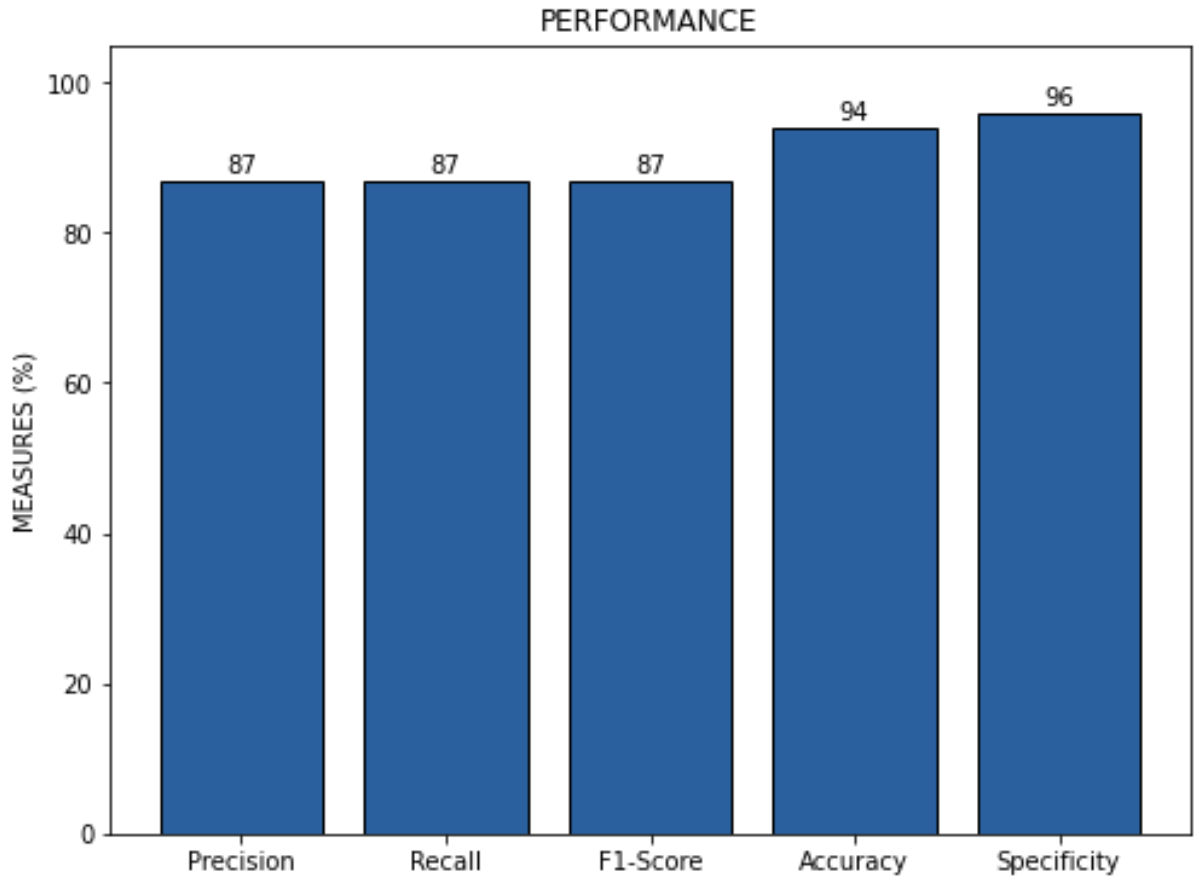
RESULTS



SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction

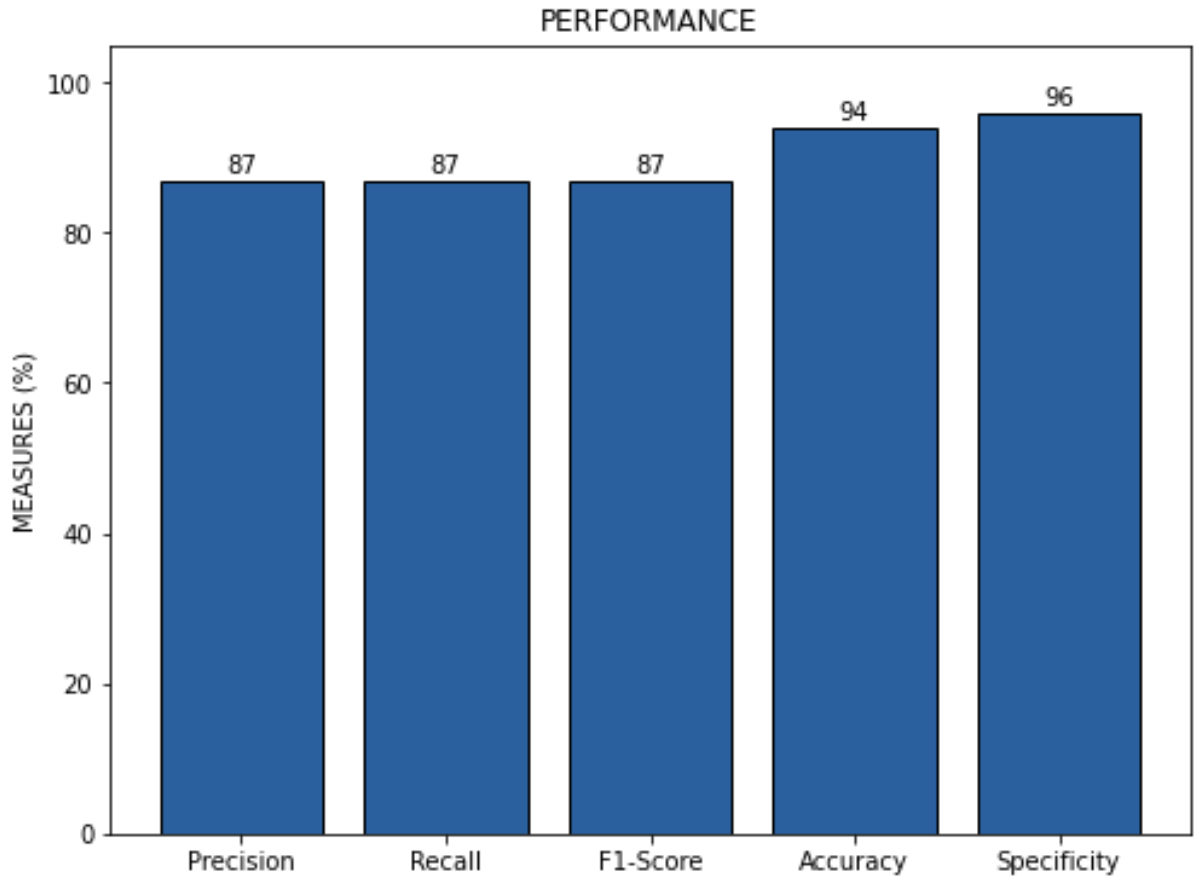
RESULTS



SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%

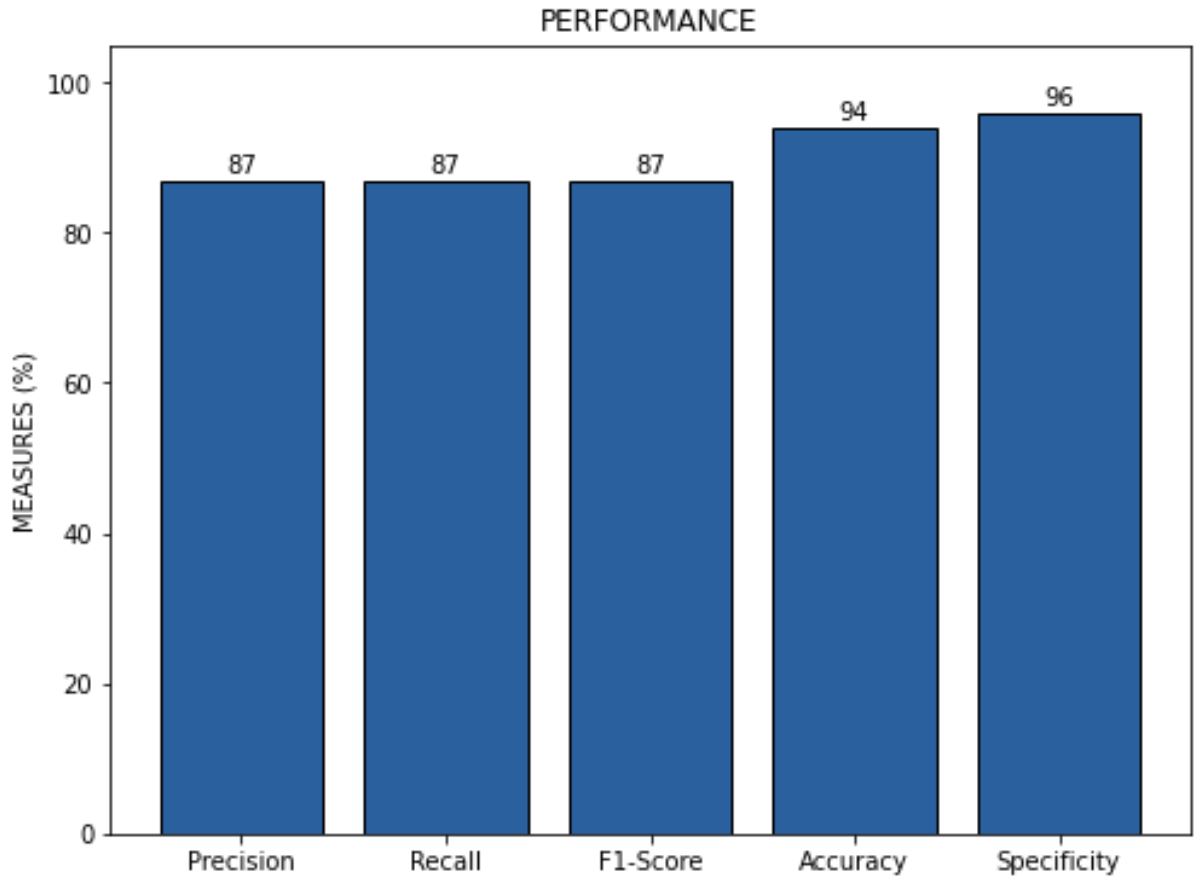
RESULTS



SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach

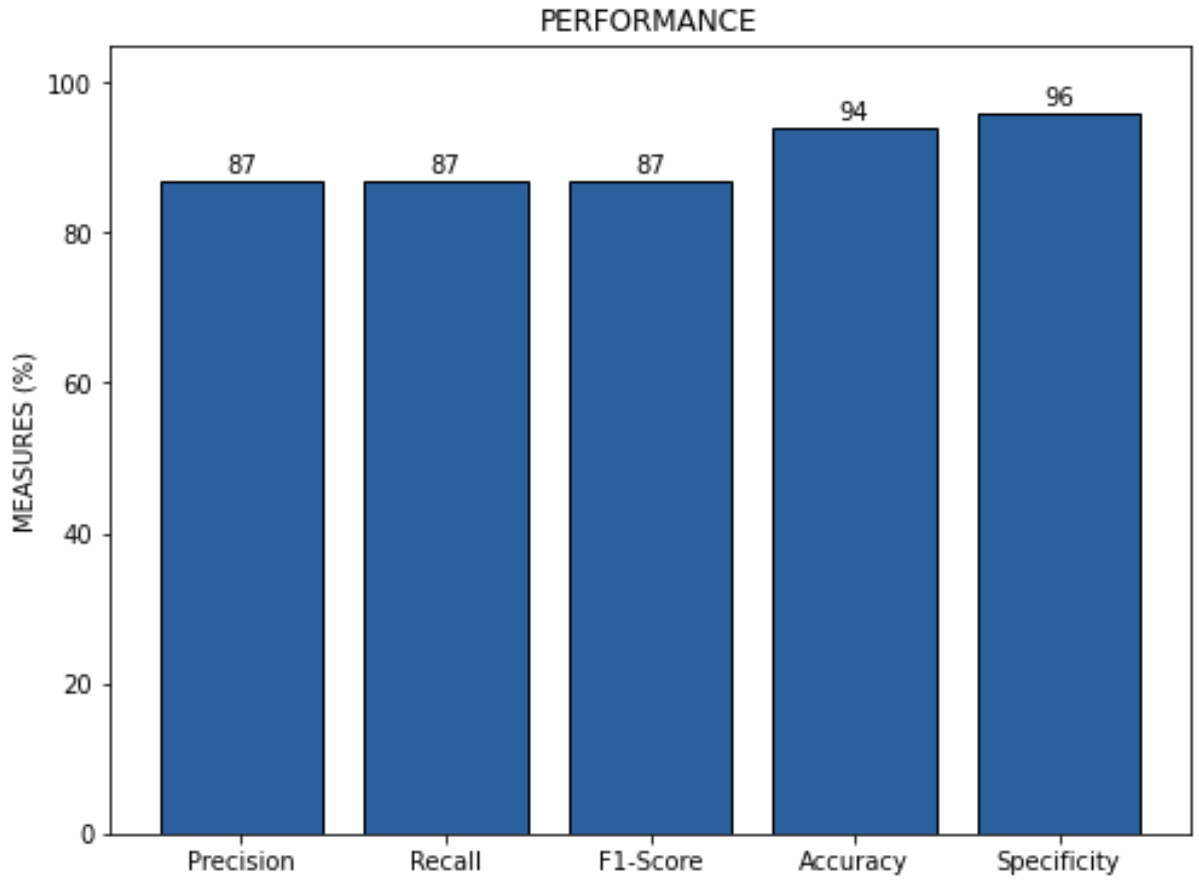
RESULTS



SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

RESULTS

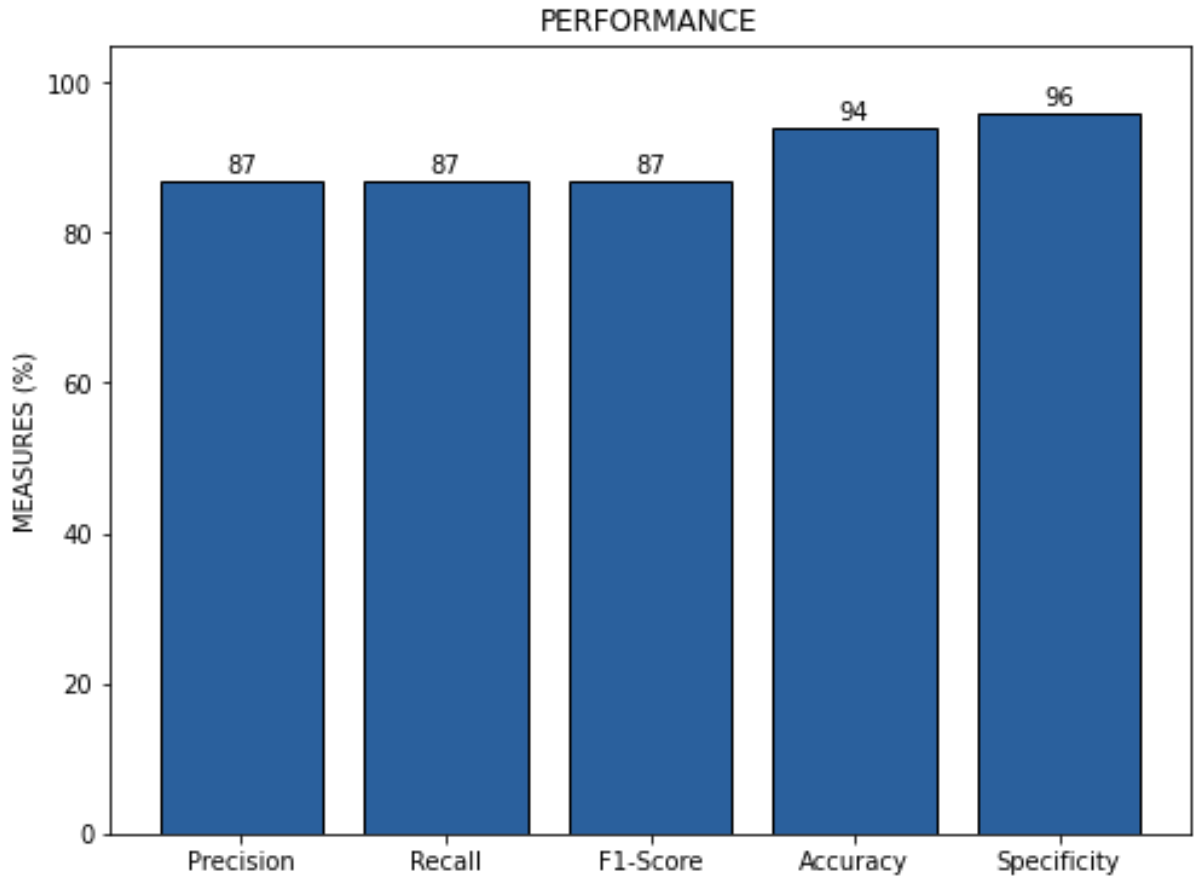


SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

COMMON ISSUES

RESULTS



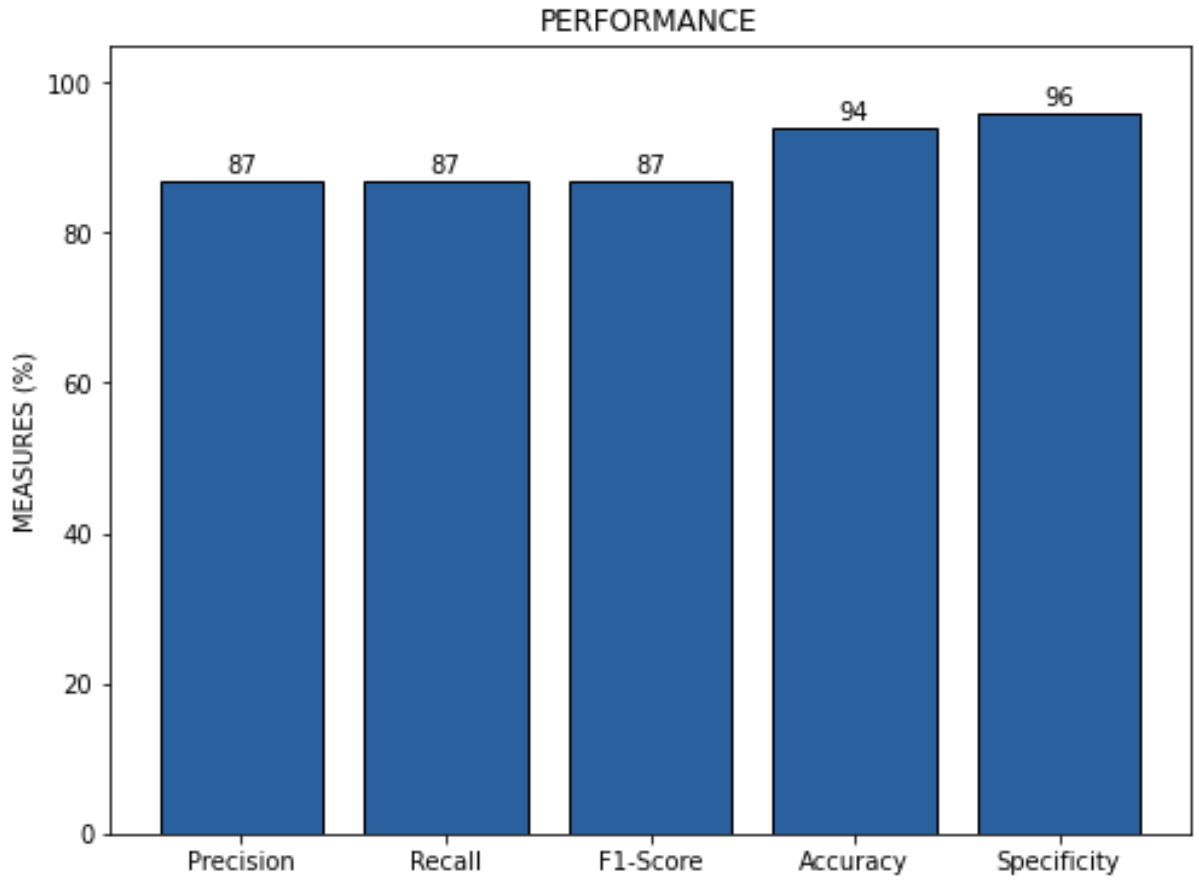
SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

COMMON ISSUES

- high false-negative rate

RESULTS



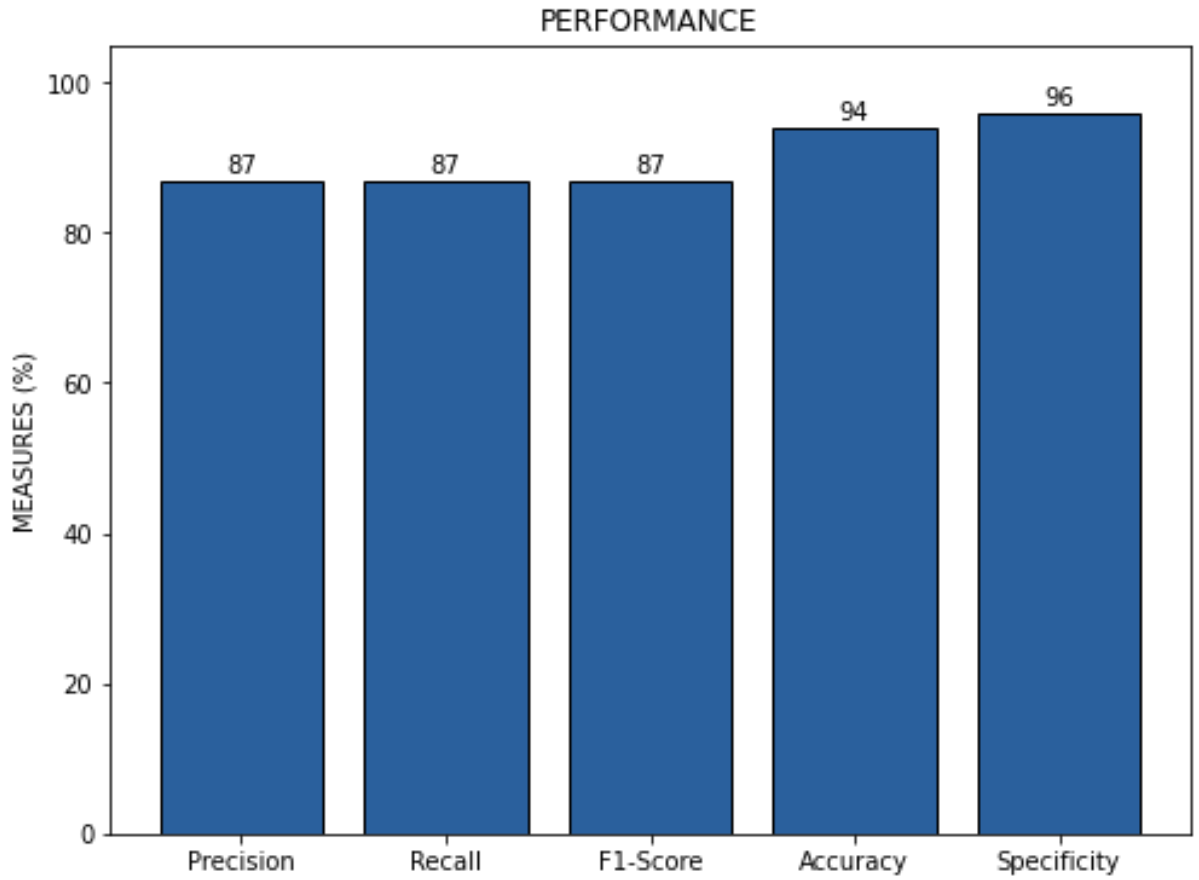
SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

COMMON ISSUES

- high false-negative rate
- detect only a few type of CTC

RESULTS



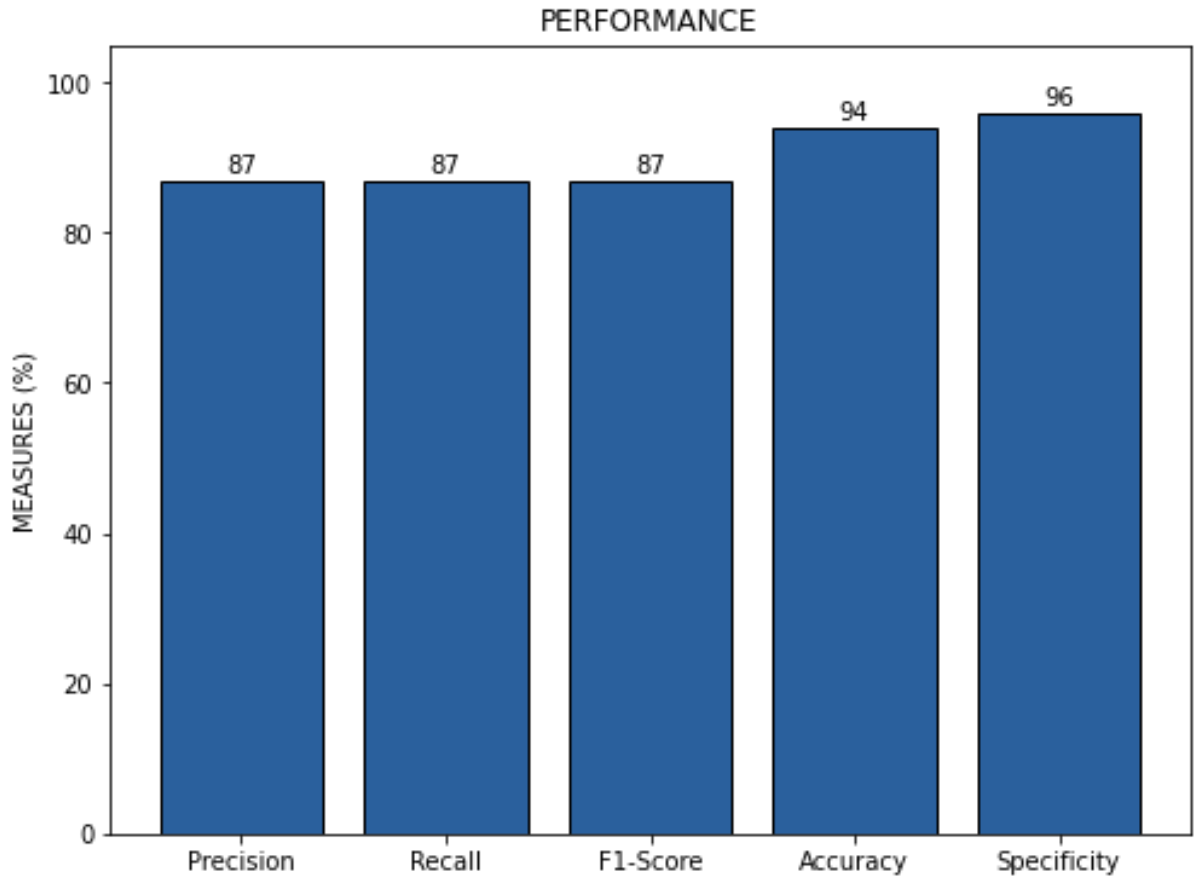
SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

COMMON ISSUES

- high false-negative rate
- detect only a few type of CTC
- not a full integrated approach

RESULTS



SOME STATE OF ART ACCURACY RESULTS

- image preprocessing, feature extraction
  - SVM, DT, NB 95%
- statistical approach
  - KNN 96%

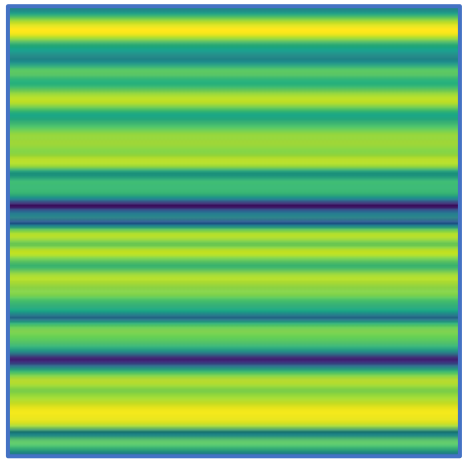
COMMON ISSUES

- high false-negative rate
- detect only a few type of CTC
- not a full integrated approach

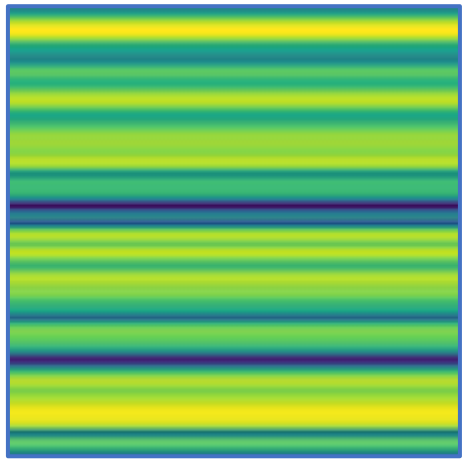
M. A. Elsadig and A. Gafar, "Covert Channel Detection: Machine Learning Approaches," in IEEE Access, vol. 10, pp. 38391-38405, 2022

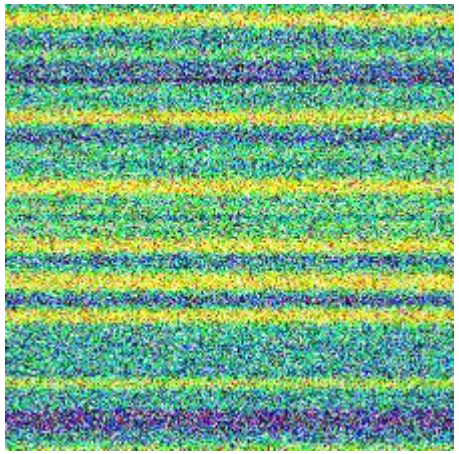
RESULTS – GAUSSIAN NOISE

RESULTS – GAUSSIAN NOISE



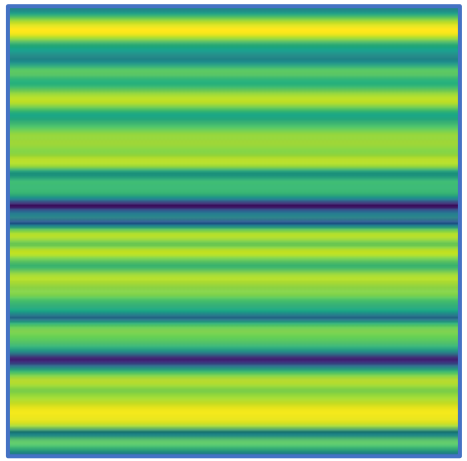
RESULTS – GAUSSIAN NOISE

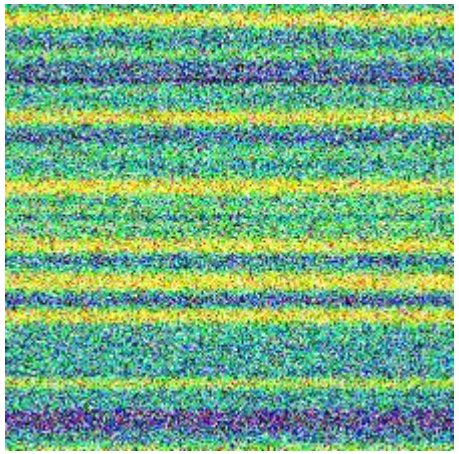




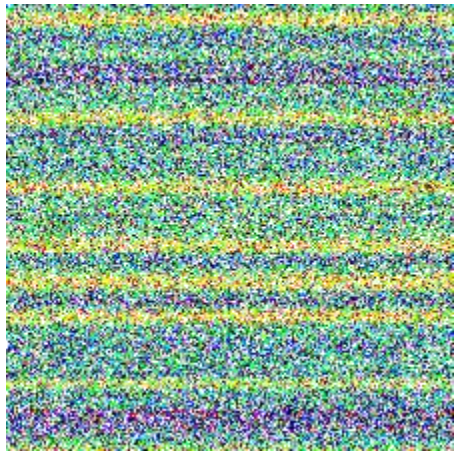
standard deviation 0.05

RESULTS – GAUSSIAN NOISE



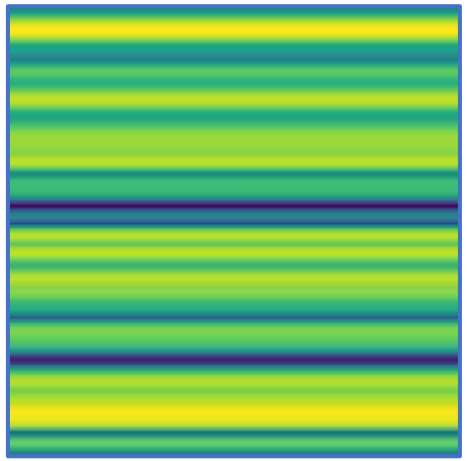


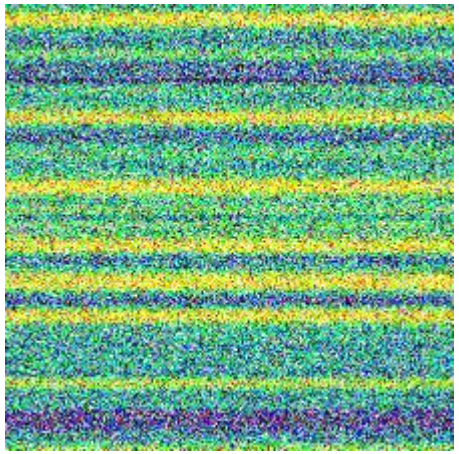
standard deviation 0.05



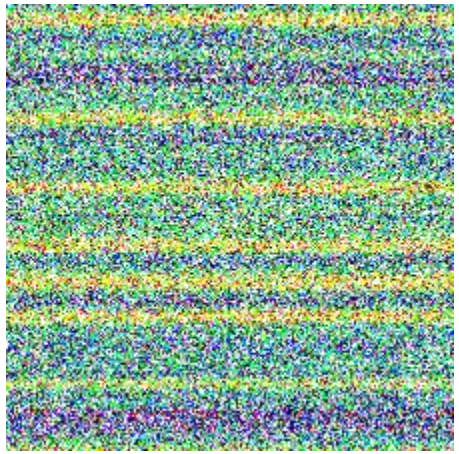
standard deviation 0.5

RESULTS – GAUSSIAN NOISE

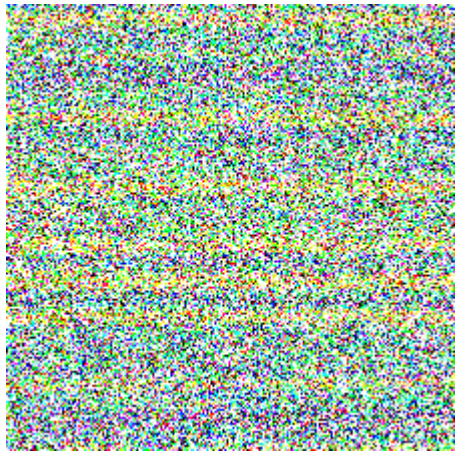




standard deviation 0.05

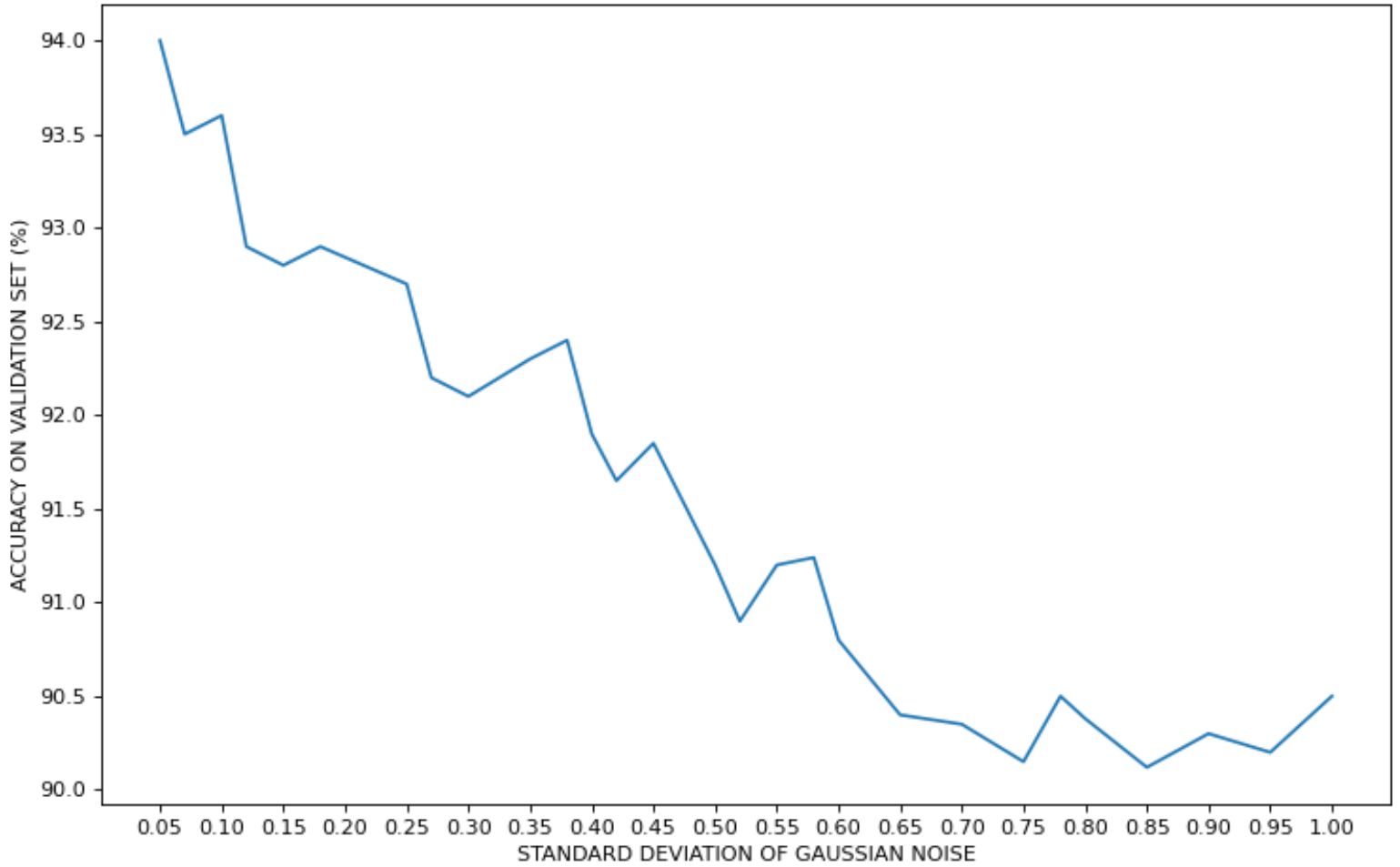


standard deviation 0.5

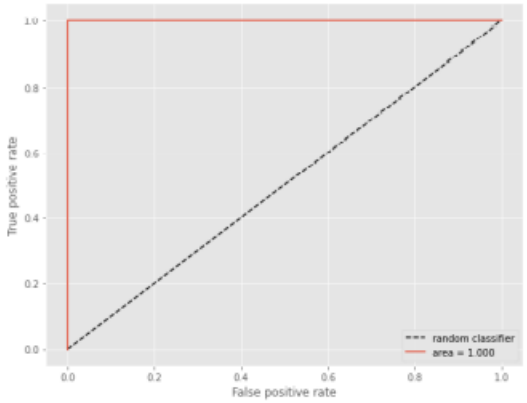


standard deviation 1.0

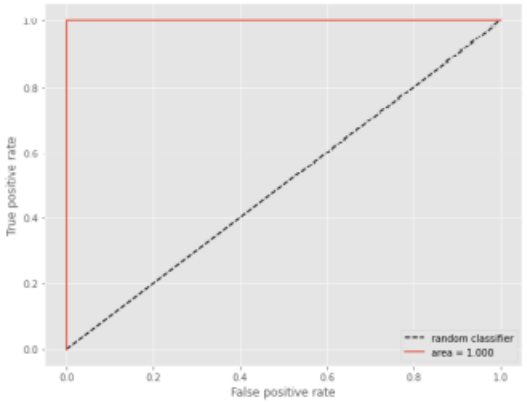
RESULTS – GAUSSIAN NOISE



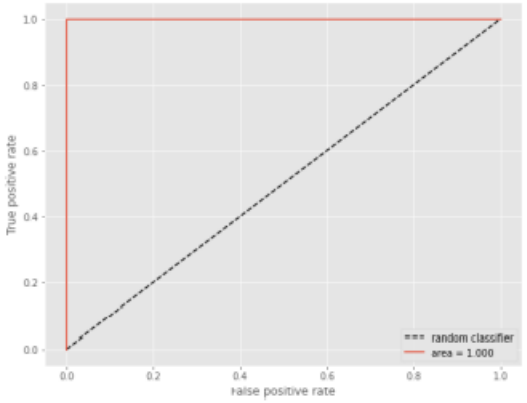
RESULTS – RECEIVER OPERATOR CHARACTERISTIC



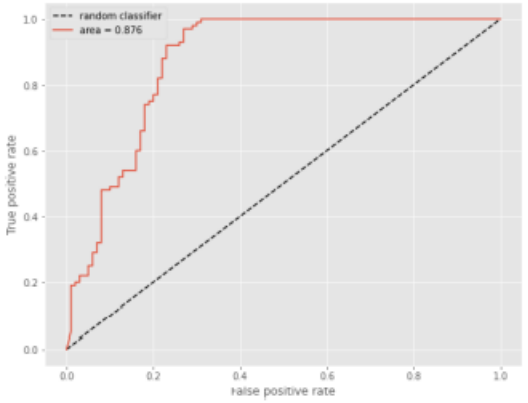
(a) ROC credit-card e email



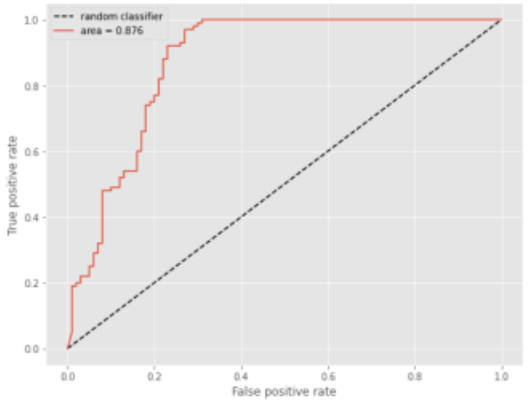
(b) ROC credit-card e vuota



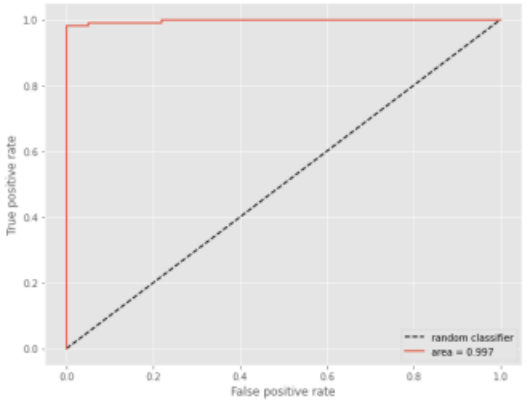
(e) ROC email e vuota



(f) ROC nome-cognome e credit-card



(c) ROC credit-card e nome-cognome



(d) ROC email e nome-cognome

CONCLUSION AND FUTURE DEVELOPMENTS

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan
  - no other communication

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan
  - no other communication

- learned while having fun! ☺

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan
  - no other communication

- learned while having fun! ☺
- submitted (i) paper for IEEE VTS *VTC2023-Spring*
  (ii) paper for ASTES journal

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan
  - no other communication

- learned while having fun! ☺
- submitted (i) paper for IEEE VTS *VTC2023-Spring*
  (ii) paper for ASTES journal

- deep packet inspection

CONCLUSION AND FUTURE DEVELOPMENTS

- robust approach against noise
- each CTC type spectrogram has a *prototype*
- safe environment
  - same lan
  - no other communication

- learned while having fun! ☺
- submitted (i) paper for IEEE VTS *VTC2023-Spring*
  (ii) paper for ASTES journal

- deep packet inspection
- other CTC types

THANKS *(english)*

THANKS *(english)*

谢谢 *(chinese)*

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

شكرا (*arabic*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

شكرا (*arabic*)

Спасибо (*russian*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

شكرا (*arabic*)

Спасибо (*russian*)

OBRIGADO (*portuguese*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

شكرا (*arabic*)

Спасибо (*russian*)

OBRIGADO (*portuguese*)

থ্যাংকয়ু (*bengali*)

THANKS (*english*)

Спасибо (*russian*)

谢谢 (*chinese*)

OBRIGADO (*portuguese*)

धन्यवाद (*hindi*)

থ্যাংকয়ু (*bengali*)

GRACIAS (*spanish*)

DANKE (*german*)

MERCI (*french*)

شكرا (*arabic*)

THANKS (*english*)

Спасибо (*russian*)

谢谢 (*chinese*)

OBRIGADO (*portuguese*)

धन्यवाद (*hindi*)

থ্যাংকয়ু (*bengali*)

GRACIAS (*spanish*)

DANKE (*german*)

MERCI (*french*)

ありがとう (*japanese*)

شكرا (*arabic*)

THANKS (*english*)

Спасибо (*russian*)

谢谢 (*chinese*)

OBRIGADO (*portuguese*)

धन्यवाद (*hindi*)

থ্যাংকয়ু (*bengali*)

GRACIAS (*spanish*)

DANKE (*german*)

MERCI (*french*)

ありがとう (*japanese*)

شكرا (*arabic*)

감사합니다 (*korean*)

THANKS (*english*)

谢谢 (*chinese*)

धन्यवाद (*hindi*)

GRACIAS (*spanish*)

MERCI (*french*)

شكرا (*arabic*)

GRAZIE!

Спасибо (*russian*)

OBRIGADO (*portuguese*)

থ্যাংকয়ু (*bengali*)

DANKE (*german*)

ありがとう (*japanese*)

감사합니다 (*korean*)