**FAISAL SHAFAIT**

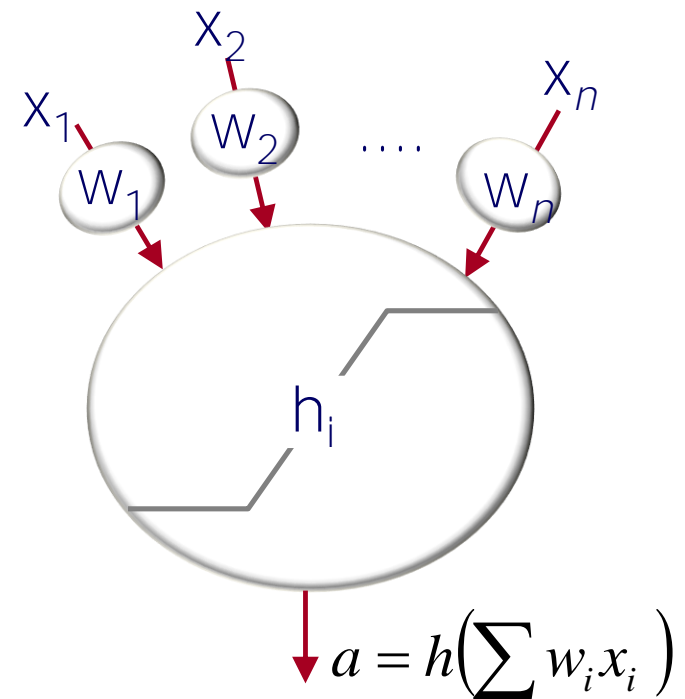# Deep Learning and
# Optical Character Recognition

# Artificial Neural Networks (ANNs)

↘ Goal: make computers "intelligent"

↘ Idea: Model human brain

*Artificial* Neural Network

Synapse

Dendrite

Cell

Nucleus

Axon

$$a = h\left(\sum w_i x_i\right)$$
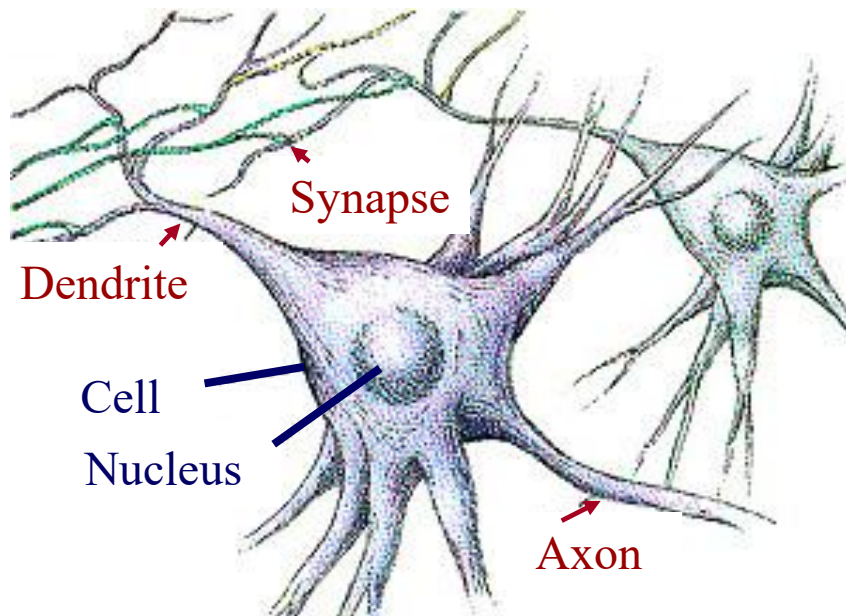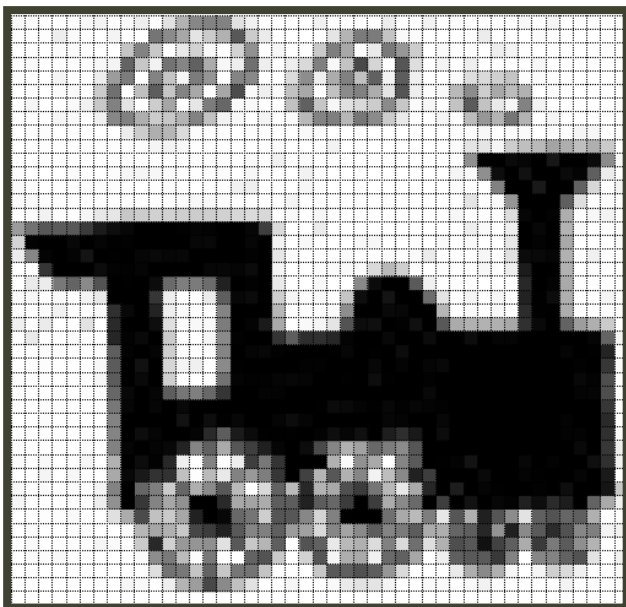
$$
\begin{array}{ccccc}
255 & 255 & 240 & \cdots & 255 \\
255 & 248 & 232 & \cdots & 255 \\
252 & 247 & 238 & \cdots & 239 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
255 & 255 & 255 & \cdots & 255
\end{array}
$$

Features

255  255  240  ⋯  255

Input Layer

Hidden Layers

$x_{i2}$
$x_{i1}$   $w_{i2}$   $x_{in}$
$w_i$   $w_{in}$

$h_i$

Output Layer

Output

0.01  ⋯  0.9  ⋯  0.2
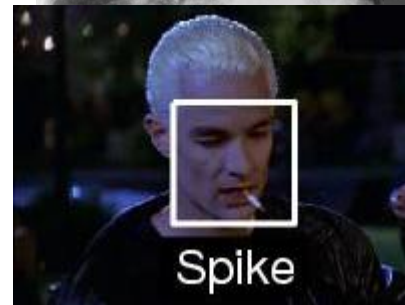
# ANN Applications: Recognize Patterns

↘ Image Analysis
- Detection (e.g., disease)
- Recognition (e.g., objects)
- Identification (e.g., persons)

↘ Data Mining
- Classification
- Change and Deviation Detection
- Knowledge Discovery

↘ Prognosis
- Ozone prognosis
- Weather Forecast
- Stock market prediction

↘ Games, …

## The Rise and Fall of ANNs

↘ ANN widely used in 1990s

↘ Suddenly went out of flavour in 2000s

↘ Renaissance – Deep Learning

↘ Popular deep architectures

- Neocognitron [Fukushima 1980]

- Recurrent Neural Networks [Hopfield 1982]

- Convolutional Neural Networks [LeCun 1989]

- Long Short-Term Memory Networks [Schmidhuber 1997]

- Deep Belief Networks [Hinton 2006]

- Self-Taught Learning [Ng 2007]

Features

Input Layer

Hidden Layer *n*

Output Layer

Output
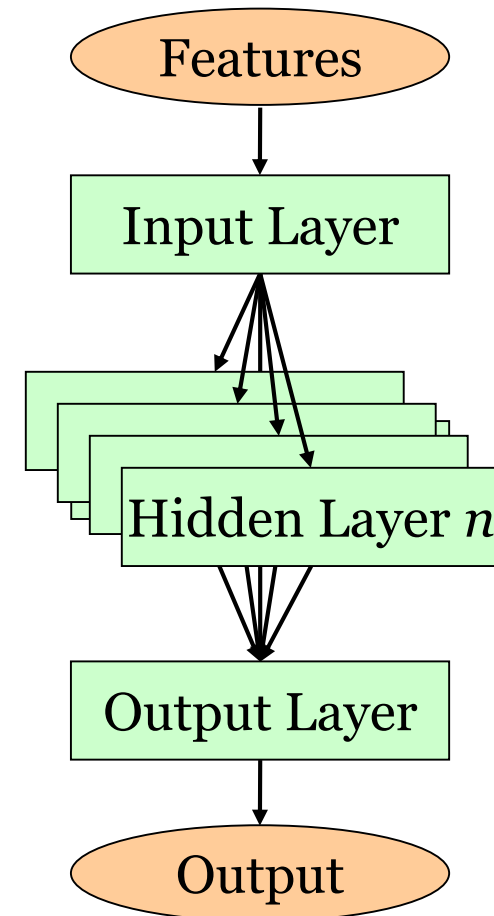
# Deep Learning Benchmarks

↘ Highest accuracy on standard benchmarks

- The MNIST Handwritten Digits Benchmark
- The NORB Object Recognition Benchmark
  The CIFAR Image Classification Benchmark

↘ Winning Competitions

- **ICDAR 2013 Arabic OCR Competition**
- MICCAI 2013 Grand Challenge on Mitosis Detection
- IJCNN 2013 Traffic Sign Recognition Contest
- ICPR 2012 contest on Mitosis Detection in Histological Images
- ISBI 2012 Brain Image Segmentation Challenge

# Deep Learning with Long Short-Term Memory (LSTM) Networks

# Recurrent Neural Networks (RNNs)

Features
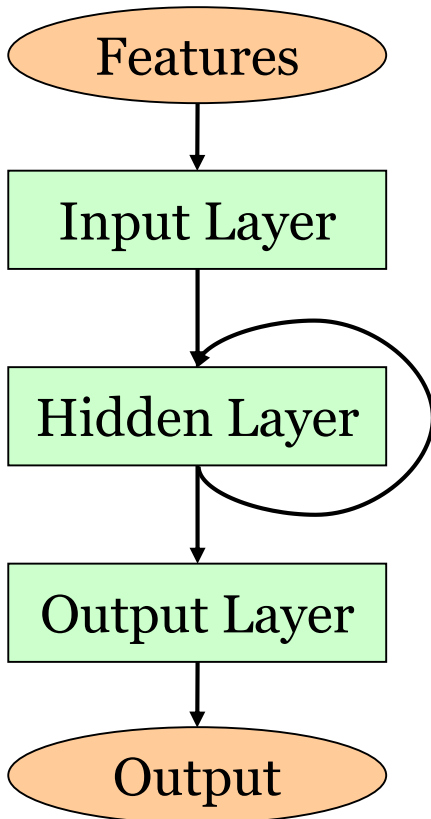
↓ Proposed by Hopfield in 1982

↓ Recurrent connections are added in order to keep information of previous time stamps in the network

↓ Novel equation for the activation:

Input Layer

Hidden Layer

$$b_h^{\,t} = h\left(\sum w_i x_i^t + \sum w_h b_h^{\,t-1}\right)$$

↓ Context information is used

↓ How to train those networks …?
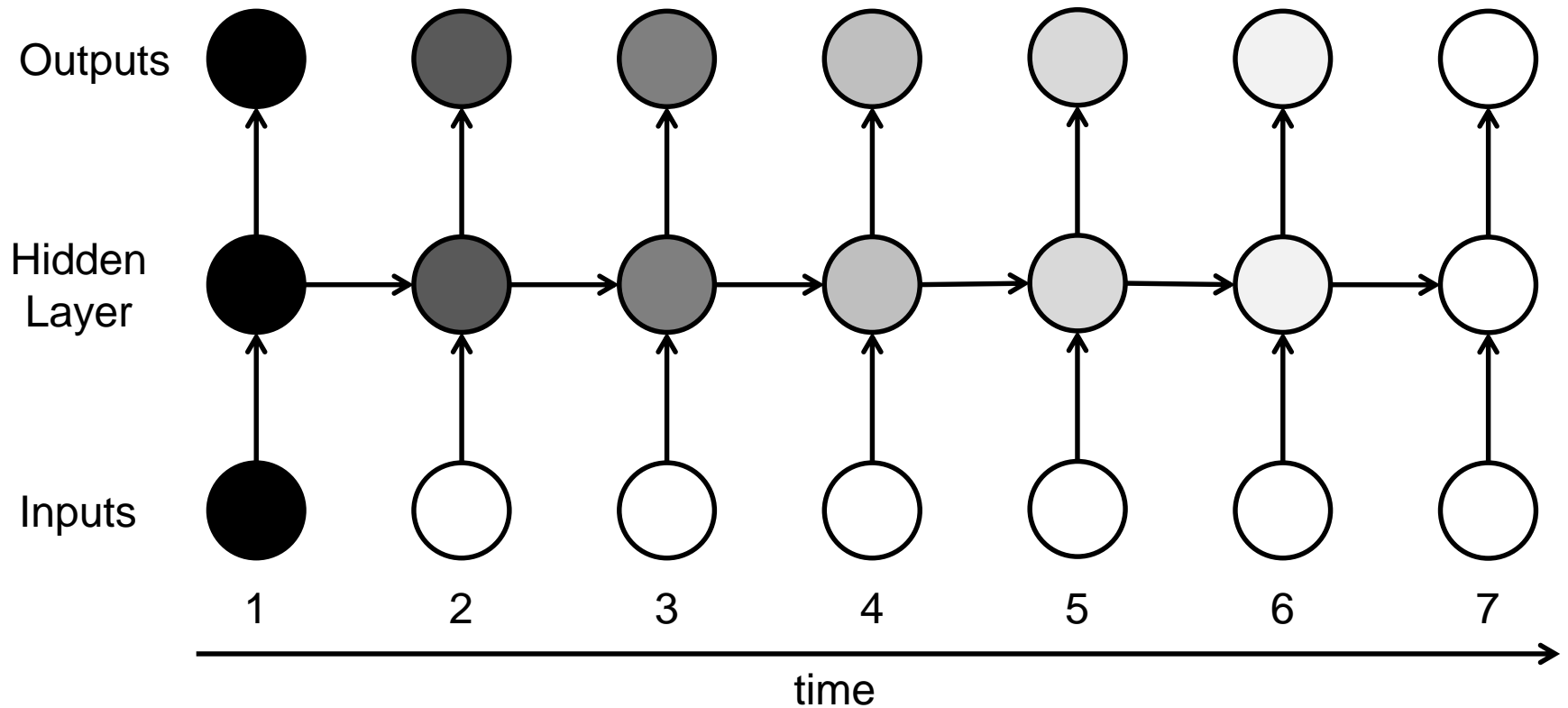
Output Layer
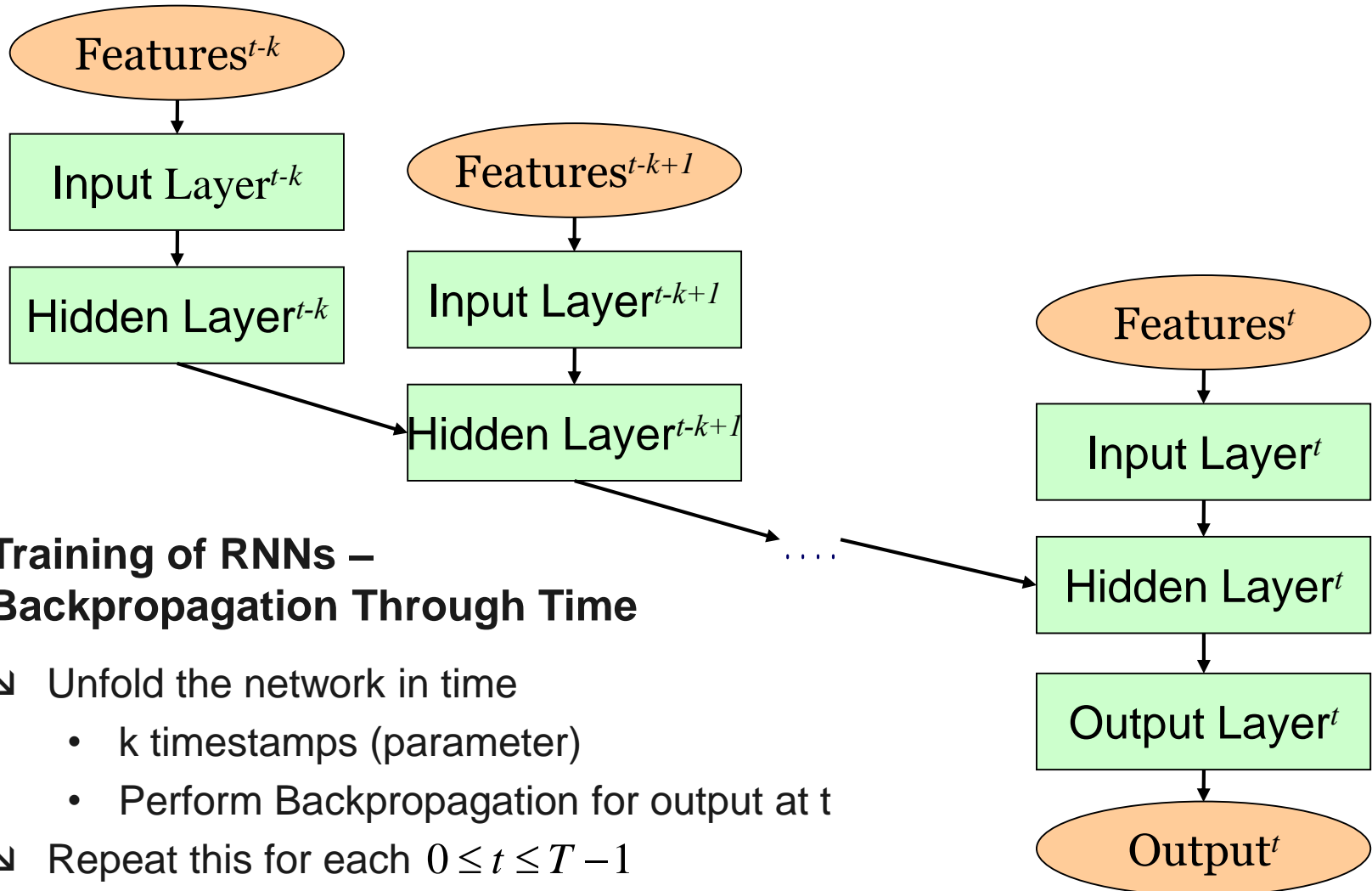
Output
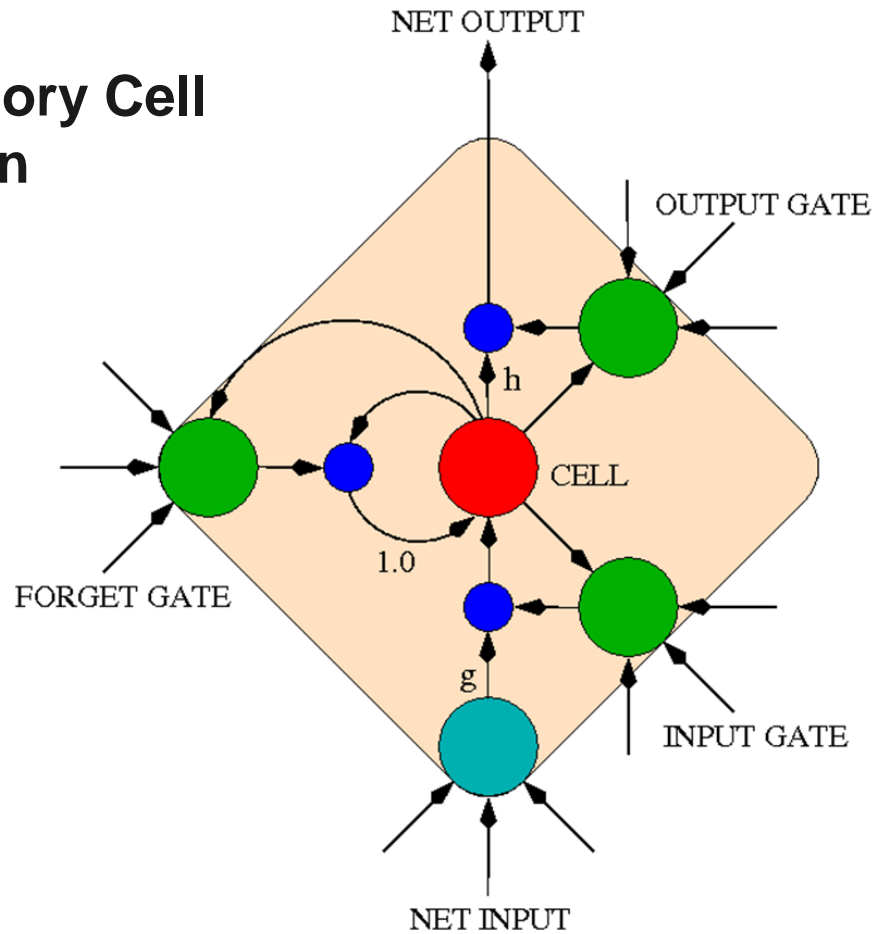
**Training of RNNs –
Backpropagation Through Time**

↘ Unfold the network in time
  • k timestamps (parameter)
  • Perform Backpropagation for output at t
↘ Repeat this for each $0 \leq t \leq T - 1$
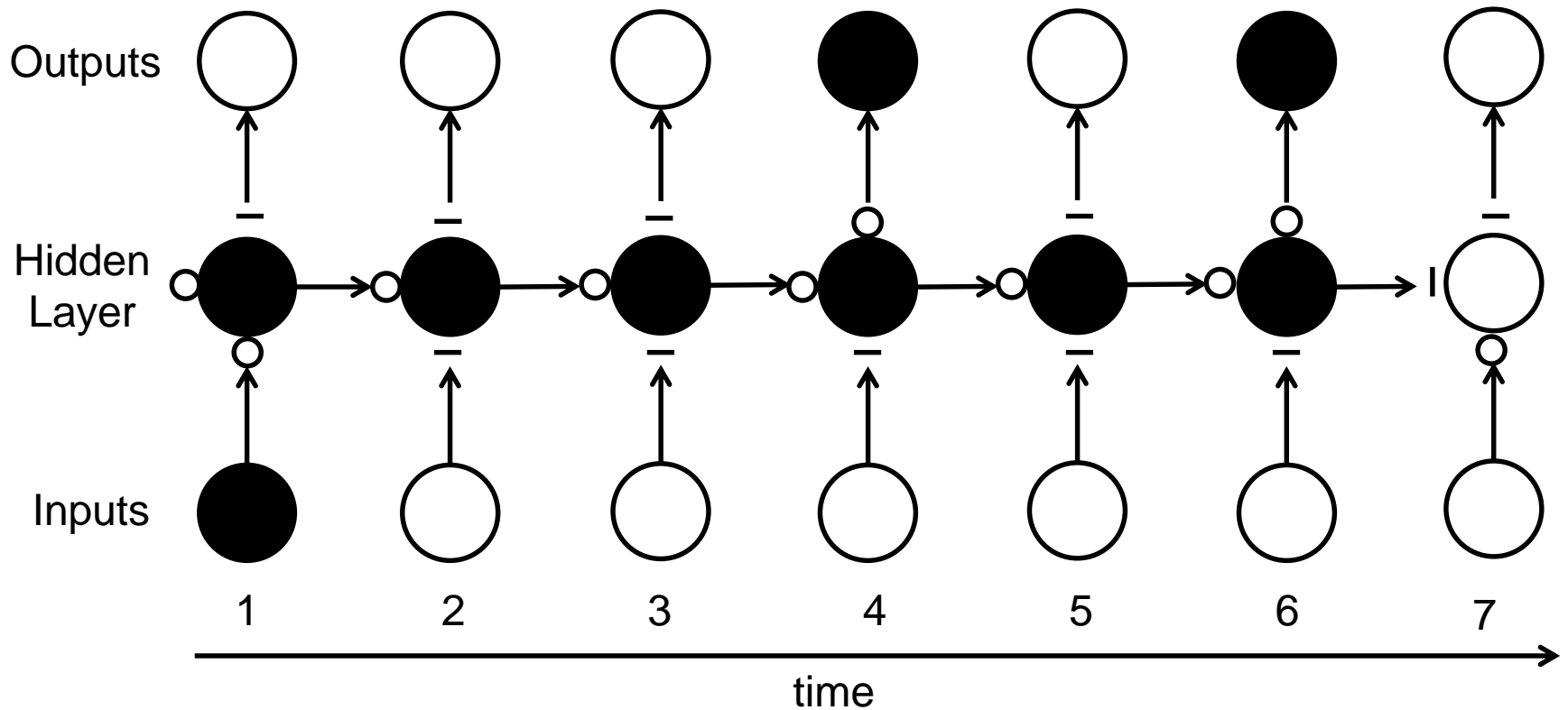
# Vanishing Gradient

## Core Idea: New Memory Cell Instead of Perceptron

## No Vanishing Gradient



Outputs

Hidden
Layer

Inputs

1    2    3    4    5    6    7

time

Input Layer

Input Gate Layer

Forget Gate Layer

Hidden Layer

Sigmoid or tanh

Multiplication

Full connection

Single connection

Output Gate Layer
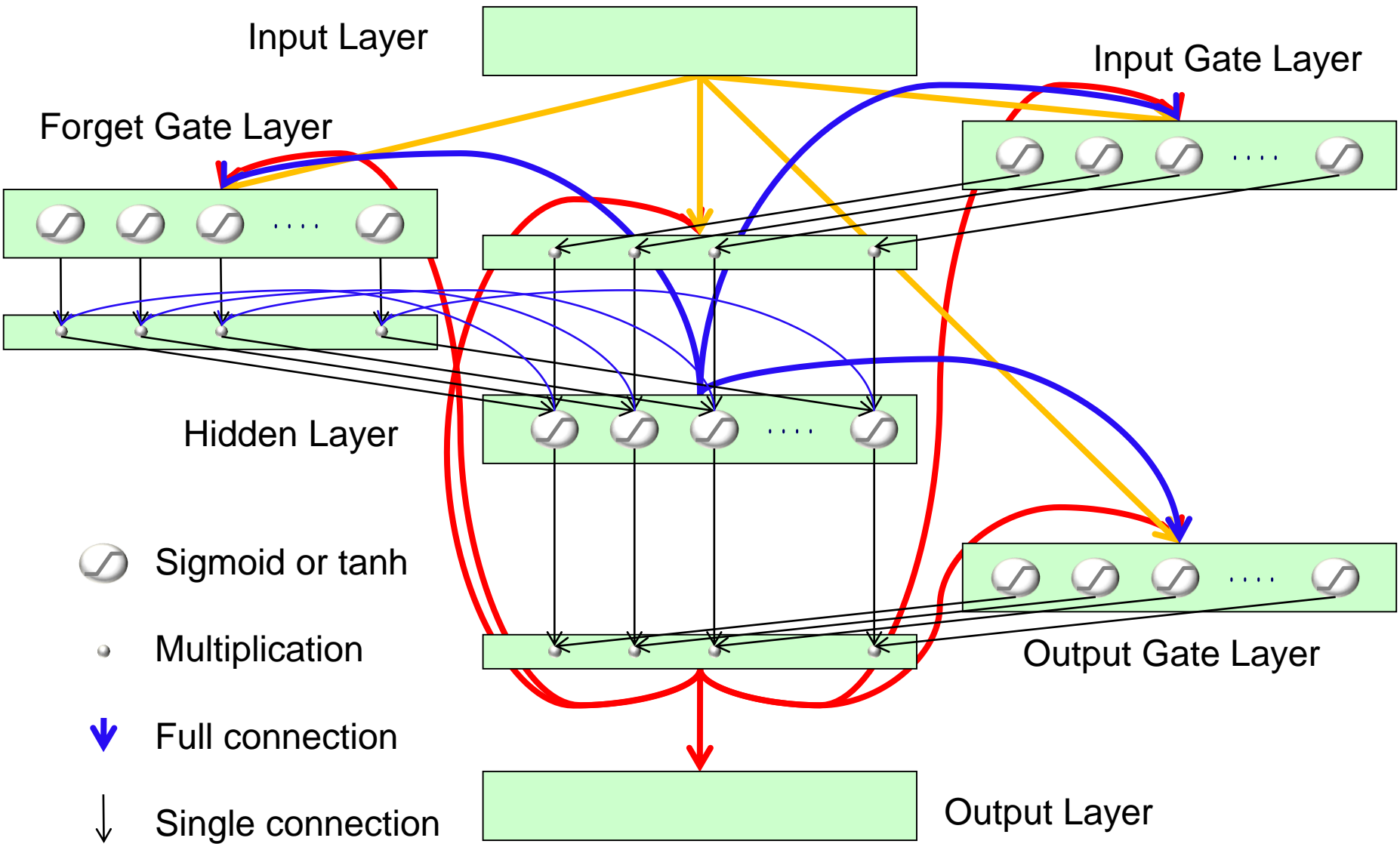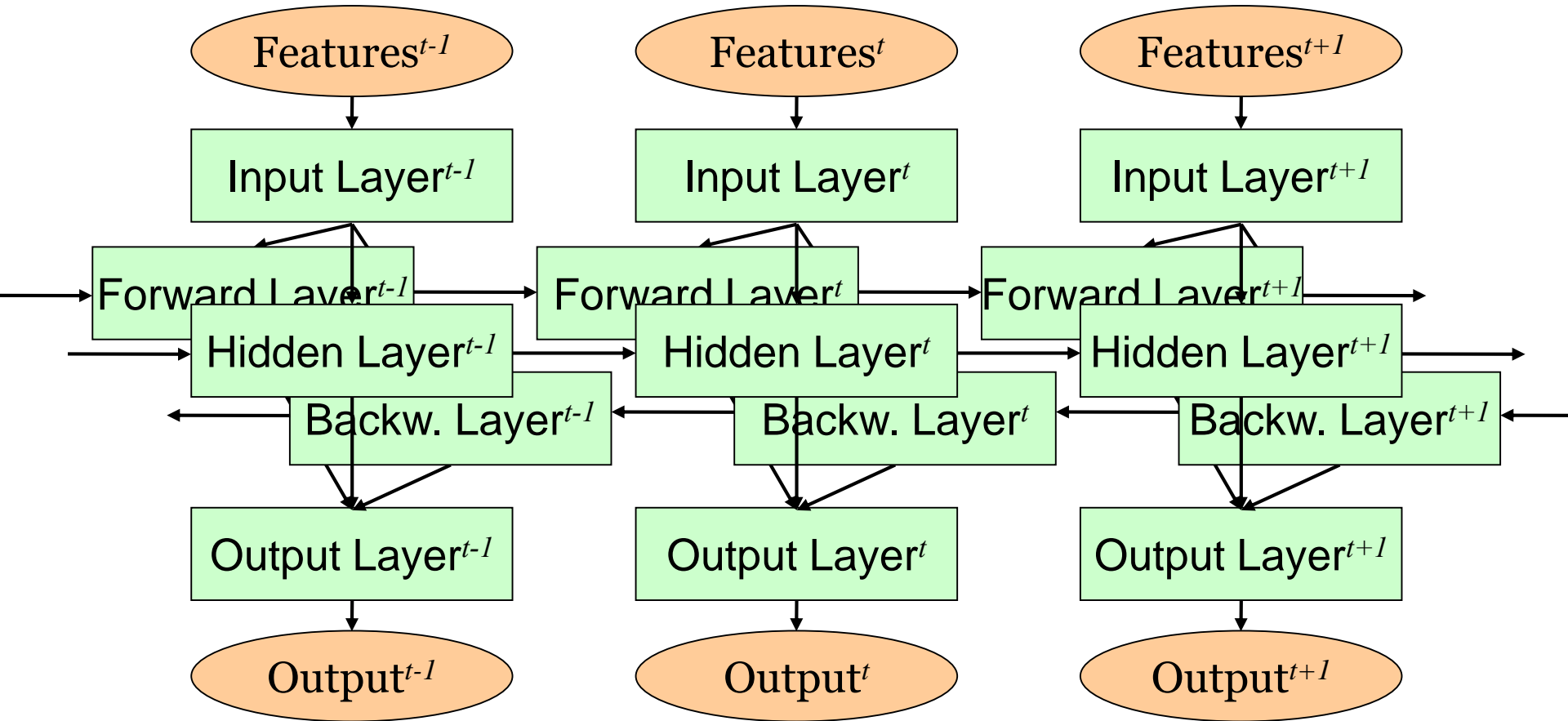
Output Layer

# Bidirectional RNN



> Trained with back-propagation through time (forward path through all time stamps for each hidden layer sequentially)

# Optical Character Recognition (OCR) with MD-BLSTM

↘ Input: raw pixel data

↘ Output machine-readable transcription

*Constitutional Irritation.* ⟶ Constitutional Irritation.

↘ Importance of context

To

Lower case "o" ?

Mathemat~~ical circle~~ symbol?

Ca~~pital "O"~~ ?

D~~igit "0"~~ ?

# Scanning Neural Network Architecture



Best Student Paper Award

**Sheikh Faisal Rashid**, **Faisal Shafait**, T Breuel. "Scanning Neural Network for Text Line Recognition", 10th IAPR Workshop on Document Analysis Systems, **DAS'12**. Gold Coast, Australia, Mar. 2012.

16

# Latin OCR with BLSTM

| OCR System | English | Fontana |
|---|---|---|
| OCRopus | 2.14 | - |
| Tesseract | 1.30 | 0.91 |
| FineReader | 0.85 | 1.23 |
| **BLSTM** | **0.59** | **0.15** |

T Breuel, **Adnan ul Hasan**, M Al-Azawi, and **Faisal Shafait**. "High-Performance OCR for Printed English and Fraktur Using LSTM Networks", 12th Int. Conf. on Document Analysis and Recognition, **ICDAR'13**. Washington DC, USA, Aug 2013.

# Urdu OCR with BLSTM

↘ Cursive script

↘ No word spacing

↘ Small inter-line gap

# Urdu OCR with MD-BLSTM

**Adnan ul Hasan**, S. Ahmed, **Sheikh Faisal Rashid**, **Faisal Shafait**, T Breuel. "Offline Printed Urdu Nastaleeq Script Recognition with Bidirectional LSTM Networks", 12th Int. Conf. on Document Analysis and Recognition, **ICDAR'13**. Washington DC, USA, **2013**.

# Urdu OCR with MD-BLSTM



Saeeda Naz, A. Umar, R. Ahmad, M. I. Razzak, **Sheikh Faisal Rashid**, **Faisal Shafait**, "Urdu Nastaliq Text Recognition using Implicit Segmentation based on Multi-Dimensional Long Short Term Memory Neural Networks", SpringerPlus, **2016**

# Results of the ICDAR 2013 Arabic OCR Contest

↘ Organized in four challenges
  1. Font (B) in 12 pt size
  2. Font (B) in multiple sizes
  3. Font (I) in multiple sizes
  4. All fonts in multiple sizes

↘ Our system (jointly developed with Siemens) won the **TOP** place in **all** four challenges with a **significant** margin

A : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

B : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

C : **نقدّم في هذا البحث قاعدة بيانات لكلمات عربية**

D : نقدم في هذا البحث قاعدة بيانات لكلمات عربية

E : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

F : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

G : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

H : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

I : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

J : نقدّم في هذا البحث قاعدة بيانات لكلمات عربية

Fouad Slimane, Slim Kanoun, Haikal El Abed, Adel M. Alimi, Rolf Ingold, Jean Hennebert: **ICDAR2013 Competition on Multi-font and Multi-size Digitally Represented Arabic Text**. 12th International Conference on Document Analysis and Recognition, ICDAR 2013: 1433-1437

# Conclusion

↘ Deep learning architectures simulate human brain

↘ During the years they became more powerful

- Better architectures and algorithms
- Faster hardware

↘ Diverse application areas

↘ Training deep architectures needs

- many CPU cores
- a lot of patience

↘ Effective training remains an art [LeCun 2013]

# Questions / Comments ?