universität freiburg

Recognition of Complex Table Structures and Extraction of Tabular data from Photovoltaic module datasheets

Master Thesis Defense

Presented by

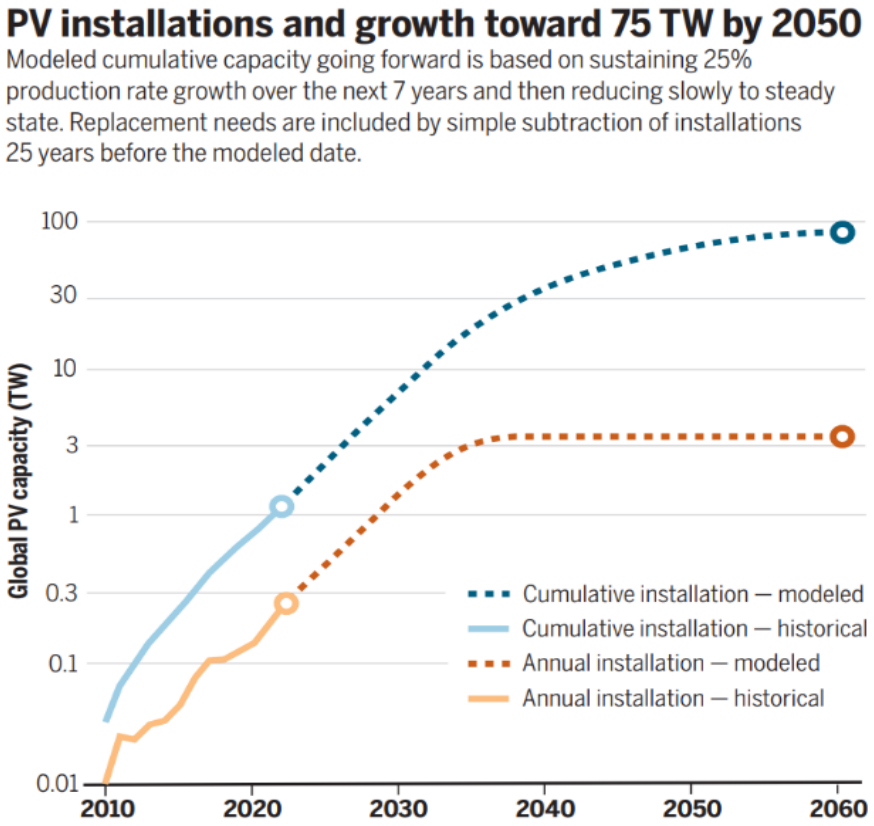
Swathi Thiruvengadam,
M.Sc. Computer Science (AI Specialisation),
Technical Faculty, University of Freiburg,
Freiburg im Breisgau, March 2025

Chapter 1
Introduction

Chapter 1 - Introduction

1.1. Motivation

- Rapid expansion of solar energy consumption
- Need accurate, automated data extraction to assist research and PV manufacturing



haegel2023

Chapter 1 - Introduction

VDS-S144/M6H-BG

VDS POWER

ELECTRICAL DATA (STC)					
Peak Power Watts-PMAX (Wp)*	430	435	440	445	450
Maximum Power Voltage-VMP (V)	40.5	40.8	41.1	41.4	41.7
Maximum Power Current-IMP (A)	10.62	10.67	10.71	10.75	10.80
Open Circuit Voltage-VOC (V)	48.7	48.9	49.1	49.3	49.5
Short Circuit Current-ISC (A)	11.20	11.29	11.37	11.45	11.53
Module Efficiency ηm (%)	19.7	20.0	20.2	20.4	20.6
Power Tolerance-PMAX (W)			0~+5		

STC: Irradiance 1000W/m², moudule temperature 25°C, AM=1.5; *Measuring tolerance: ±3%

Electrical characteristics with different rear side power gain (reference to 435 Wp front)					
Peak Power-PMAX (Wp)*	457	479	500	522	544
Maximum Power Voltage-VMP (V)	40.8	40.8	40.8	40.8	40.8
Maximum Power Current-IMP (A)	11.20	11.74	12.27	12.80	13.34
Open Circuit Voltage-VOC (V)	49.0	49.1	49.2	49.3	49.4
Short Circuit Current-ISC (A)	11.80	12.36	12.93	13.49	14.05
Pmax gain	5%	10%	15%	20%	25%

STC: Power Bifaciality: 70±5%

ELECTRICAL DATA (NMOT)					
Maximum Power-PMAX (Wp)*	325	329	333	337	341
Maximum Power Voltage-VMP (V)	38.2	38.5	38.8	39.0	39.1
Maximum Power Current-IMP (A)	8.51	8.55	8.58	8.63	8.71
Open Circuit Voltage-VOC (V)	46.0	46.2	46.4	46.6	46.7
Short Circuit Current-ISC (A)	9.02	9.05	9.08	9.12	9.15

NMOT: Irradiance 800W/m², module temperature 20°C, AM=1.5, wind speed 1m/s

MECHANICAL DATA	
Solar Cells	Monocrystalline silicon 166 mm (9BB)
Cell Orientation	144 cells (6 x 24)
Module Dimensions	2095x1039x30 mm
Weight	28.5 kg
Front Glass	2.0 mm, High Transmission, AR Coated Heat Strengthened Glass
Encapsulant Material	POE/EVA
Back Glass	2.0 mm, Heat Strengthened Glass (White Grid Glass)
Frame	30 mm Anodized Aluminium Alloy
Junction Box	IP 68 rated
Cables	Photovoltaic Technology Cable 4.0 mm² Cable length 350 mm or customized length

*Please refer to regional datasheet for specied connector.

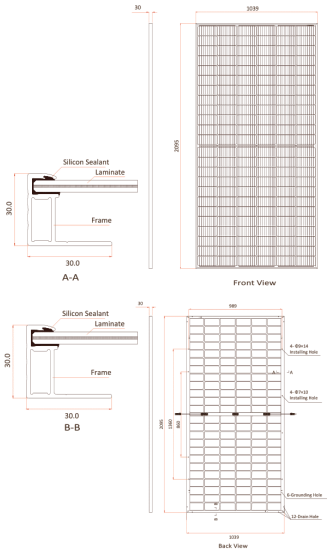
TEMPERATURE RATINGS	
NMOT (Nominal Module Operating Temperature)	41°C (±3°C)
Temperature Coefficient of PMAX	-0.34%/°C
Temperature Coefficient of VOC	-0.25%/°C
Temperature Coefficient of ISC	0.040%/°C

(Do not connect Fuse in Combiner Box with two or more strings in parallel connection)

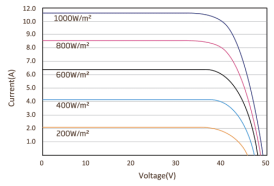
MAXIMUM RATINGS	
Operational Temperature	-40~+85°C
Maximum System Voltage	1500V DC (IEC)
Max Series Fuse Rating	20A

PACKAGING CONFIGURATION	
Modules per box	35 pieces
Modules per 40'container	770 pieces

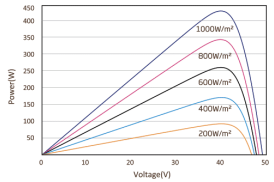
DIMENSIONS OF PV MODULE (mm)



I-V CURVES OF PV MODULE (440W)



P-V CURVES OF PV MODULE (440W)



Example of a photovoltaic module Datasheet

	A	B	C	D	E	F	G	H	I	J	K	L
1		name	year	length	width	E/eff	E/pmpp	E/vmpp	E/impp	E/voc	E/isc	EN/pmpp
2	0	5-VDS	2023	2095	1039	19.7	430	40.5	10.62	48.7	11.20	325.0
3	1	5-VDS	2023	2095	1039	20.0	435	40.8	10.67	48.9	11.29	329.0
4	2	5-VDS	2023	2095	1039	20.2	440	41.1	10.71	49.1	11.37	333.0
5	3	5-VDS	2023	2095	1039	20.4	445	41.4	10.75	49.3	11.45	337.0
6	4	5-VDS	2023	2095	1039	20.6	450	41.7	10.80	49.5	11.53	341.0

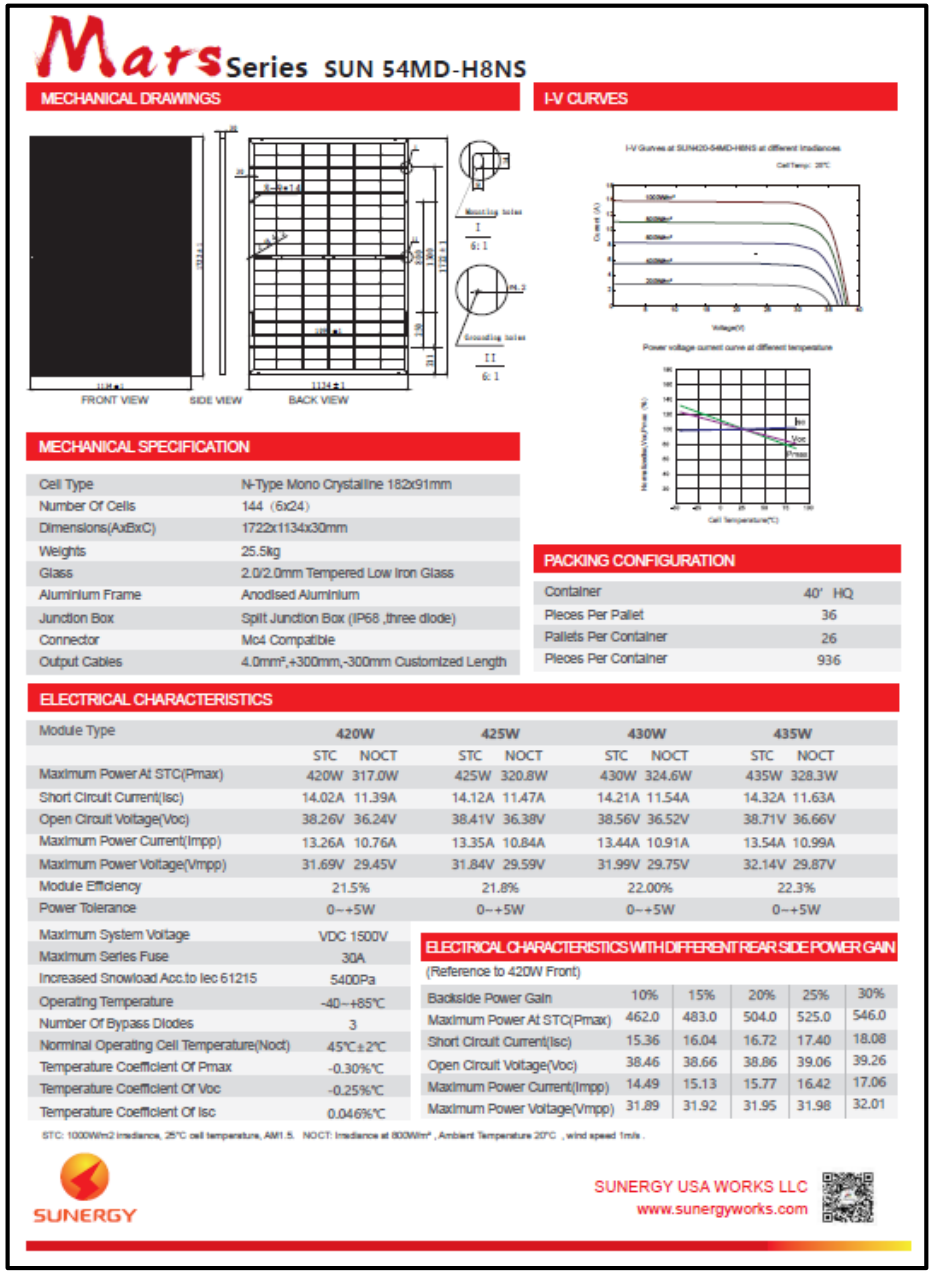
M	N	O	P	Q	R	S	T	U
EN/vmpp	EN/impp	EN/voc	EN/isc	T/isc	T/pmpp	T/voc	/pcs_palle	/pallet_con
38.2	8.51	46.0	9.02	0.040	-0.34%/°	-0.25%/°	35	770
38.5	8.55	46.2	9.03	0.040	-0.34%/°	-0.25%/°	35	770
38.8	8.58	46.4	9.08	0.040	-0.34%/°	-0.25%/°	35	770
39.0	8.63	46.6	9.12	0.040	-0.34%/°	-0.25%/°	35	770
39.1	8.71	46.7	9.15	0.040	-0.34%/°	-0.25%/°	35	770

Data extracted from the photovoltaic module datasheet

Chapter 1 - Introduction

1.2. Problems

- Unstructured, inconsistent table layouts with merged cells and multi-level headers
- Manual extraction - time-consuming, prone to errors
- Existing pipeline struggles with complex table structures and image-based PDFs



Mars Series SUN 54MD-H8NS

MECHANICAL DRAWINGS

I-V CURVES

MECHANICAL SPECIFICATION

Cell Type	N-Type Mono Crystalline 182x91mm
Number Of Cells	144 (6x24)
Dimensions(AxBxC)	1722x1134x30mm
Weights	25.5kg
Glass	2.0/2.0mm Tempered Low Iron Glass
Aluminium Frame	Anodised Aluminium
Junction Box	Split Junction Box (IP68 ,three diode)
Connector	Mc4 Compatible
Output Cables	4.0mm²,+300mm,-300mm Customized Length

PACKING CONFIGURATION

Container	40' HQ
Pieces Per Pallet	36
Pallets Per Container	26
Pieces Per Container	936

ELECTRICAL CHARACTERISTICS

Module Type	420W		425W		430W		435W	
	STC	NOCT	STC	NOCT	STC	NOCT	STC	NOCT
Maximum Power At STC(Pmax)	420W	317.0W	425W	320.8W	430W	324.6W	435W	328.3W
Short Circuit Current(Isc)	14.02A	11.39A	14.12A	11.47A	14.21A	11.54A	14.32A	11.63A
Open Circuit Voltage(Voc)	38.26V	36.24V	38.41V	36.38V	38.56V	36.52V	38.71V	36.66V
Maximum Power Current(Impp)	13.26A	10.76A	13.35A	10.84A	13.44A	10.91A	13.54A	10.99A
Maximum Power Voltage(Vmpp)	31.69V	29.45V	31.84V	29.59V	31.99V	29.75V	32.14V	29.87V
Module Efficiency	21.5%		21.8%		22.00%		22.3%	
Power Tolerance	0~+5W		0~+5W		0~+5W		0~+5W	

Maximum System Voltage	VDC 1500V
Maximum Series Fuse	30A
Increased Snowload Acc.to Iec 61215	5400Pa
Operating Temperature	-40~+85℃
Number Of Bypass Diodes	3
Norminal Operating Cell Temperature(Noct)	45℃±2℃
Temperature Coefficient Of Pmax	-0.30%℃
Temperature Coefficient Of Voc	-0.25%℃
Temperature Coefficient Of Isc	0.046%℃

ELECTRICAL CHARACTERISTICS WITH DIFFERENT REAR SIDE POWER GAIN

(Reference to 420W Front)

Backside Power Gain	10%	15%	20%	25%	30%
Maximum Power At STC(Pmax)	462.0	483.0	504.0	525.0	546.0
Short Circuit Current(Isc)	15.36	16.04	16.72	17.40	18.08
Open Circuit Voltage(Voc)	38.46	38.66	38.86	39.06	39.26
Maximum Power Current(Impp)	14.49	15.13	15.77	16.42	17.06
Maximum Power Voltage(Vmpp)	31.89	31.92	31.95	31.98	32.01

STC: 1000W/m2 Irradiance, 25°C cell temperature, AM1.5. NOCT: Irradiance at 800W/m², Ambient Temperature 20°C , wind speed 1m/s .

SUNERGY

SUNERGY USA WORKS LLC
www.sunergyworks.com

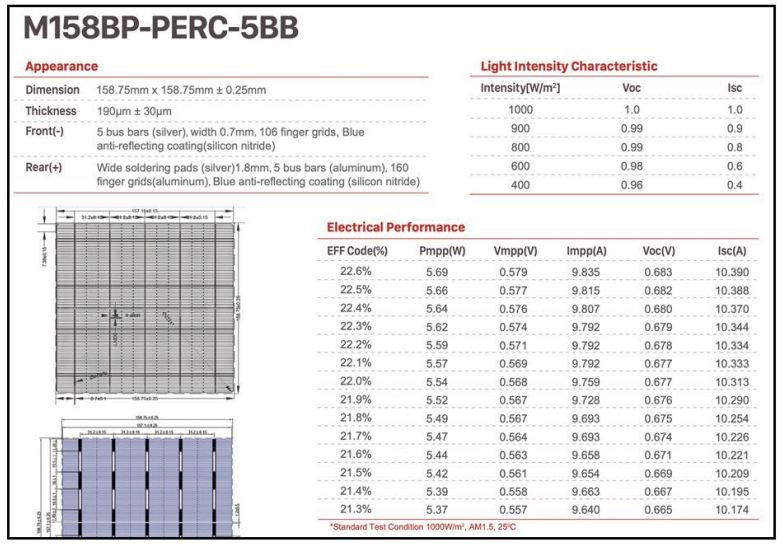
SUN-54MD-H8NS

Chapter 1 - Introduction

1.3. Current Approaches of PV data extraction

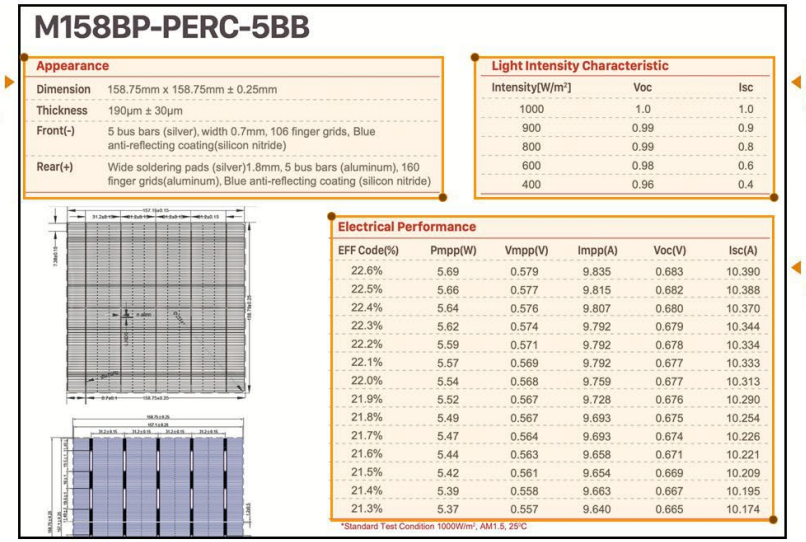
1.3.1 Lightning-Table approach overview

- DL-based detectors to identify table
- Table extraction using Tabula / Camelot
- Maps extracted data to known categories



PDF Document

Table Detection



PDF Document with Tables Detected

Malik2023

Chapter 1 - Introduction

1.3. Current Approaches of PV data extraction

1.3.1 Lightning-Table approach overview

- DL-based detectors to identify table
- Table extraction using Tabula / Camelot
- Maps extracted data to known categories

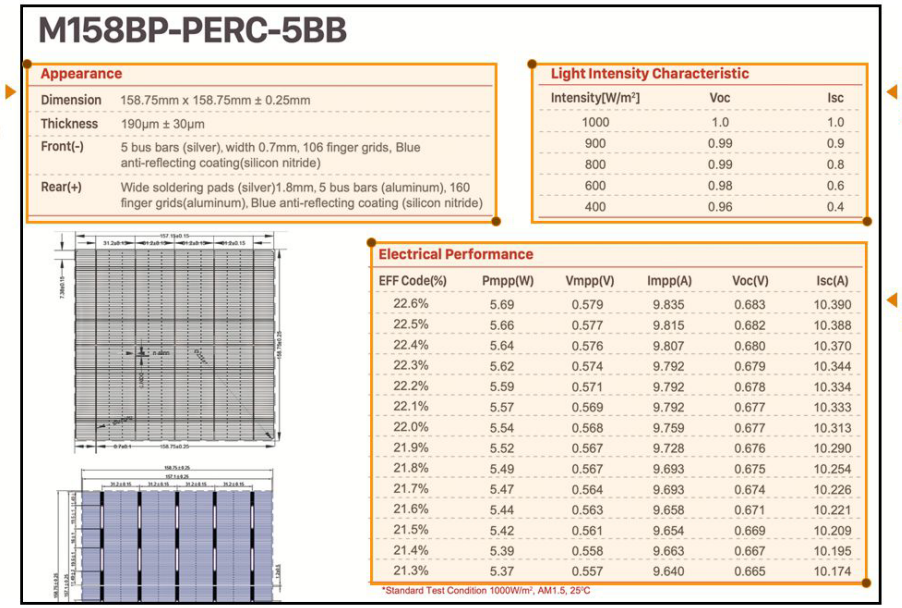


Table Recognition

	A	B	C	D	E	F
1	EFF Code(%	Pmpp(W)	Vmpp(V)	Impp(A)	Voc(V)	Isc(A)
2	22.6%	5.69	0.579	9.835	0.683	10.390
3	22.5%	5.66	0.577	9.815	0.682	10.388
4	22.4%	5.64	0.576	9.807	0.680	10.370
5	22.3%	5.62	0.574	9.792	0.679	10.344
6	22.2%	5.59	0.571	9.792	0.678	10.334
7	22.1%	5.57	0.569	9.792	0.677	10.333
8	22.0%	5.54	0.568	9.759	0.677	10.313
9	21.9%	5.52	0.567	9.728	0.676	10.290
10	21.8%	5.49	0.567	9.693	0.675	10.254
11	21.7%	5.47	0.564	9.693	0.674	10.226
12	21.6%	5.44	0.563	9.658	0.671	10.221
13	21.5%	5.42	0.561	9.654	0.669	10.209
14	21.4%	5.39	0.558	9.663	0.667	10.195
15	21.3%	5.37	0.557	9.640	0.665	10.174

Raw Values Extracted from Tables

Malik2023

Chapter 1 - Introduction

1.3. Current Approaches of PV data extraction

1.3.1 Lightning-Table approach overview

- DL-based detectors to identify table
- Table extraction using Tabula / Camelot
- Maps extracted data to known categories

	A	B	C	D	E	F
1	EFF Code(%	Pmpp(W)	Vmpp(V)	Impp(A)	Voc(V)	Isc(A)
2	22.6%	5.69	0.579	9.835	0.683	10.390
3	22.5%	5.66	0.577	9.815	0.682	10.388
4	22.4%	5.64	0.576	9.807	0.680	10.370
5	22.3%	5.62	0.574	9.792	0.679	10.344
6	22.2%	5.59	0.571	9.792	0.678	10.334
7	22.1%	5.57	0.569	9.792	0.677	10.333
8	22.0%	5.54	0.568	9.759	0.677	10.313
9	21.9%	5.52	0.567	9.728	0.676	10.290
10	21.8%	5.49	0.567	9.693	0.675	10.254
11	21.7%	5.47	0.564	9.693	0.674	10.226
12	21.6%	5.44	0.563	9.658	0.671	10.221
13	21.5%	5.42	0.561	9.654	0.669	10.209
14	21.4%	5.39	0.558	9.663	0.667	10.195
15	21.3%	5.37	0.557	9.640	0.665	10.174

Raw Values Extracted from Tables

Final Step

Efficiency = [`22.6%`, `22.5%`, `22.4%`, ... , `21.3%`]

Power @ Max Power Point = [`5.69`, `5.66`, `5.64`, ... , `5.37`]

.
.
.

Validated & Structured Values

Malik2023

Chapter 1 - Introduction

1.3.2 Lightning-Table Drawbacks

Electrical Specifications				
Module Type	GH400M6-B66HS/BT-C	GH405M6-B66HS/BT-C	GH410M6-B66HS/BT-C	GH415M6-B66HS/BT-C
Testing Condition	STC	STC	STC	STC
Pmax(W)	400	405	410	415
Imp(A)	10.77	10.86	10.95	11.04
Vmp(V)	37.17	37.33	37.48	37.63
Isc(A)	11.20	11.30	11.40	11.50
Voc(V)	45.67	45.82	45.97	46.12
Module Efficiency	17.86%	18.08%	18.31%	18.53%
		Bifaciality:70%±10%	STC:AM1.5 1000W/㎡ 25℃	Measurement uncertainty:±3%

Example of a table

	A	B	C	D	E	F	G	H
1	Electrical Specifications							
2	Module Type	GH400M6-B66HS/BT-C	GH405M6-B66HS/BT-C	GH410M6-B66HS/BT-C	GH415M6-B66HS/BT-C			
3	Testing Condition	STC	STC (a)	STCSTC				
4	Pmax(W) (b)	400	405	410415				
5	Imp(A)	10.77	10.86					
6				10.9511.04				
7	Vmp(V)	37.17						
8			37.33	37.4837.63				
9	Isc(A)	11.20	11.30	11.4011.50				
10	Voc(V)		45.82	45.9746.12				
11		45.67						
12	Module Efficiency	17.86%	18.08%	18.31%18.53%				
13	(c)			Bifaciality:70%±10%	STC:AM1.5 1000W/㎡ 25℃	Measurement uncertainty:±3%		

Data extracted using Lightning-Table

Chapter 1 - Introduction

1.3.2 Lightning-Table Drawbacks

Electrical Parameters

Module Type	SPICN6(LAR)-60-375/IH		SPICN6(LAR)-60-380/IH		SPICN6(LAR)-60-385/IH	
	STC	NOCT	STC	NOCT	STC	NOCT
Maximum Power (Pmax/W)	375	282	380	286	385	290
Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40.0
Short Circuit Current (Isc/A)	11.50	9.23	11.61	9.32	11.72	9.42
Module Efficiency	20.9%		21.2%		21.5%	

Example of a table

	A	B	C	D	E	F	G	H	I	J
1	Module Type		SPICN6(LAR)-60- 375/IH			SPIC N6(LAR)-60- 380/IH			SPICN6(LAR)- 60-385/IH	
2										
3		STC		NOCT	STC		NOCT	STC		NOCT
4										
5	Maximum Power (Priax/W)	375		282	380		286	385		290
6										
7	Maximum Power Voltage (Vmpp/V)	35.1		32.5	35.3		32.7	25.5		32.9
8										
9	Maximum Power Current (Impp/A)	10.69		8.68	10.77		875	10.85		8.82
10										
11	Open Circuit, Voltage (Voc/V)	41.6		39.8	417		39.9	41.8		40.0
12										
13	Short Circuit Current (Isc/A)	11.50		9.23	11.61		9.32	172		9.42
14										
15	Module Efficiency		20.9%			21.2%			21.5%	
16										

Data extracted using Lightning-Table

Chapter 1 - Introduction

1.4. Thesis Objectives

1. **Objective 1**: Develop an automated DL pipeline for detecting tables, recognizing its structure and extracting complex tabular data from PV module datasheets

2. **Objective 2**: Enhance table recognition performance for various table layouts (horizontal, vertical, dual-axis)

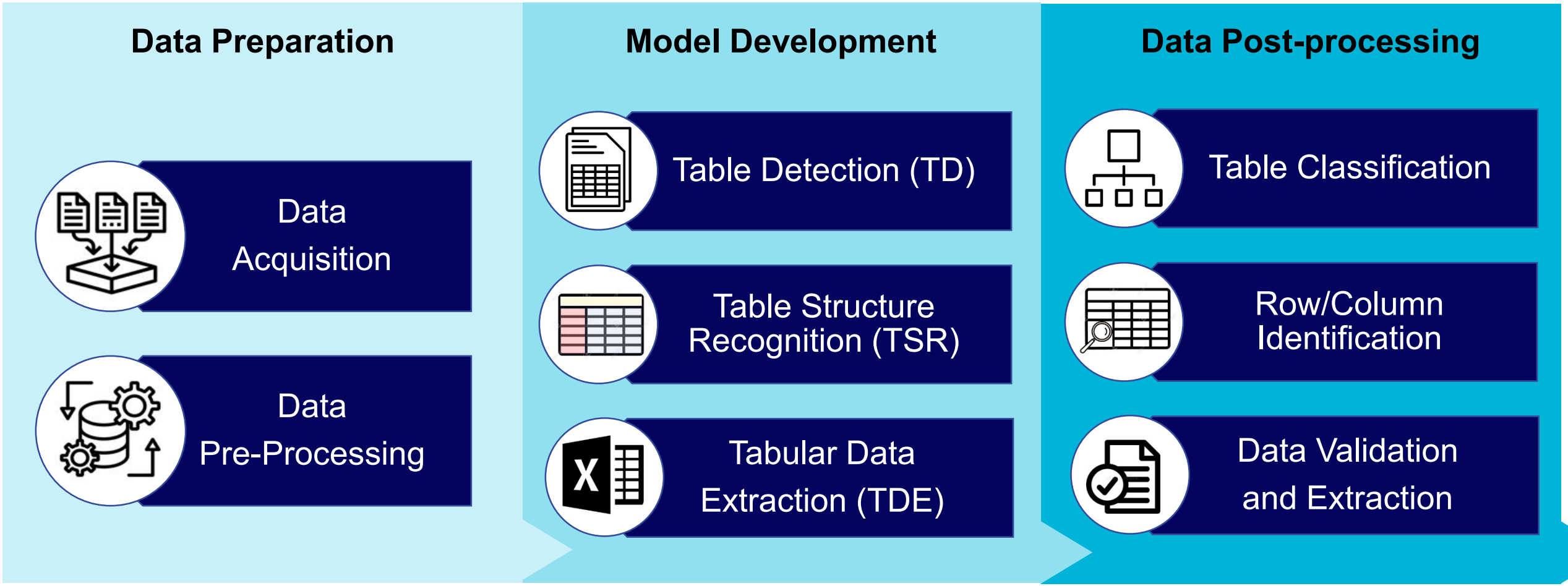
3. **Objective 3**: Integrate post-processing techniques to validate and structure key data for advanced analysis

Information Extraction

Chapter 2
Approach and Results

Chapter 2 – Approach and Results

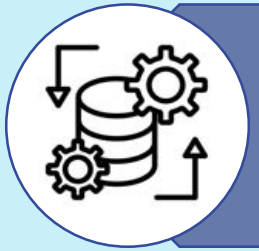
2.0. Thesis overview



Chapter 2 – Approach and Results

Data Preparation

Data Acquisition



Data Pre-Processing

2.1. Data Acquisition

2.1.1. Data gathering and annotation

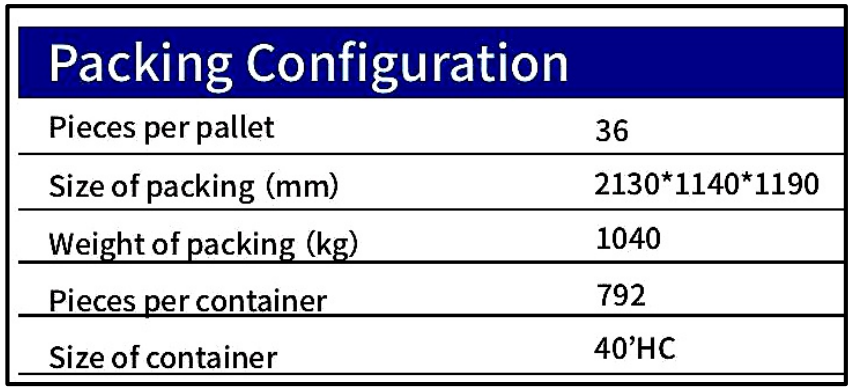
- Publicly Available PV Datasheets from manufacturers website
- Manual Annotation using LabelImg

2.1.2. Data Augmentation

- Increasing Dataset size and variety
- Resolves Class Imbalance

Chapter 2 – Approach and Results

2.1.1. Data gathering and annotation

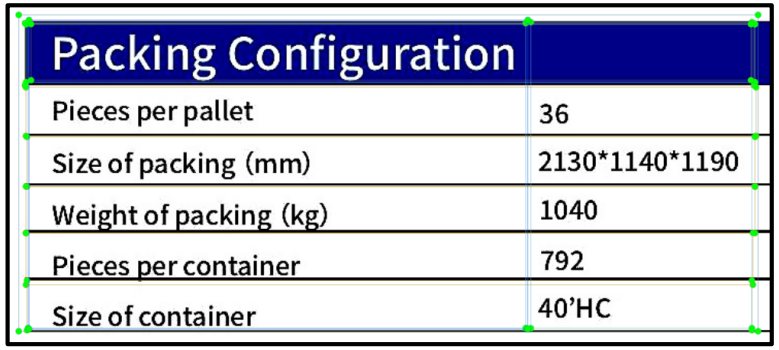


Packing Configuration	
Pieces per pallet	36
Size of packing (mm)	2130*1140*1190
Weight of packing (kg)	1040
Pieces per container	792
Size of container	40'HC

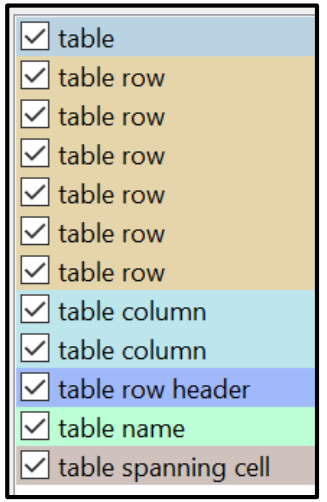
An example of input image

LabelImg

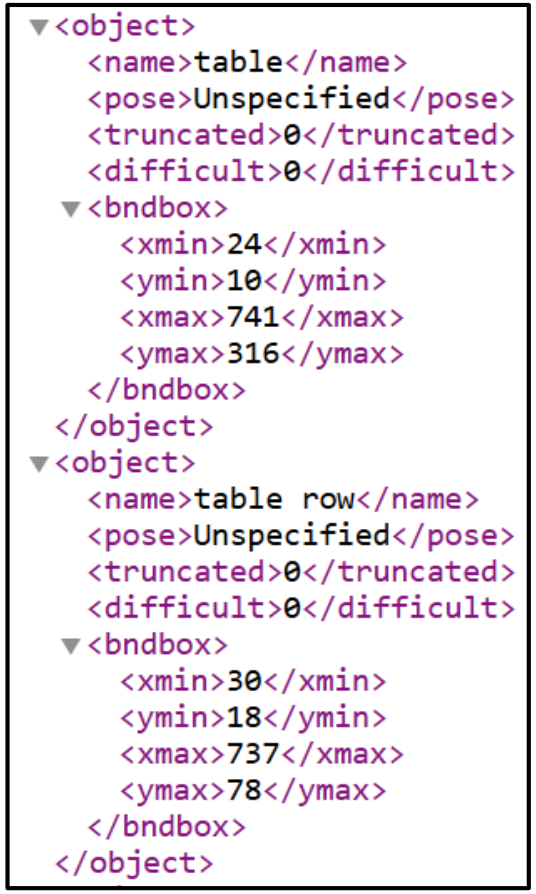
Annotation tool



Packing Configuration	
Pieces per pallet	36
Size of packing (mm)	2130*1140*1190
Weight of packing (kg)	1040
Pieces per container	792
Size of container	40'HC



An example of annotated image with class labels

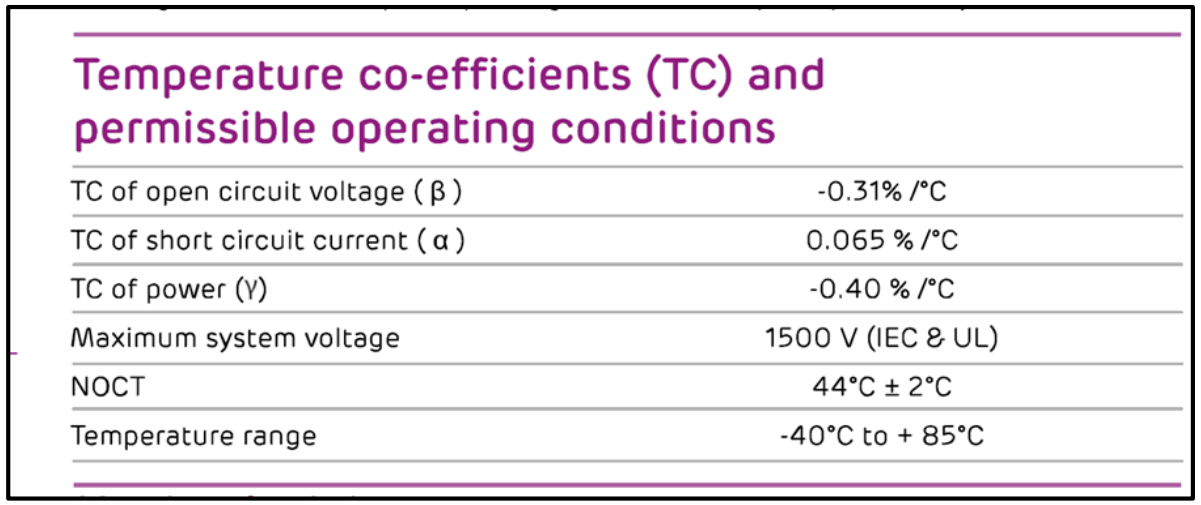


```
<object>
  <name>table</name>
  <pose>Unspecified</pose>
  <truncated>0</truncated>
  <difficult>0</difficult>
  <bndbox>
    <xmin>24</xmin>
    <ymin>10</ymin>
    <xmax>741</xmax>
    <ymax>316</ymax>
  </bndbox>
</object>
<object>
  <name>table row</name>
  <pose>Unspecified</pose>
  <truncated>0</truncated>
  <difficult>0</difficult>
  <bndbox>
    <xmin>30</xmin>
    <ymin>18</ymin>
    <xmax>737</xmax>
    <ymax>78</ymax>
  </bndbox>
</object>
```

Position annotations

Chapter 2 – Approach and Results

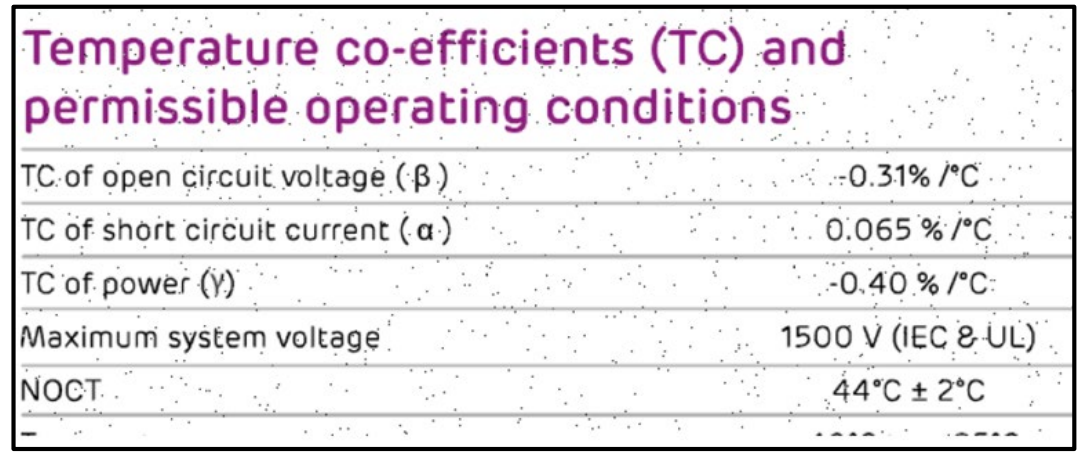
2.1.2. Data Augmentation



Temperature co-efficients (TC) and permissible operating conditions

TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C
Temperature range	-40°C to + 85°C

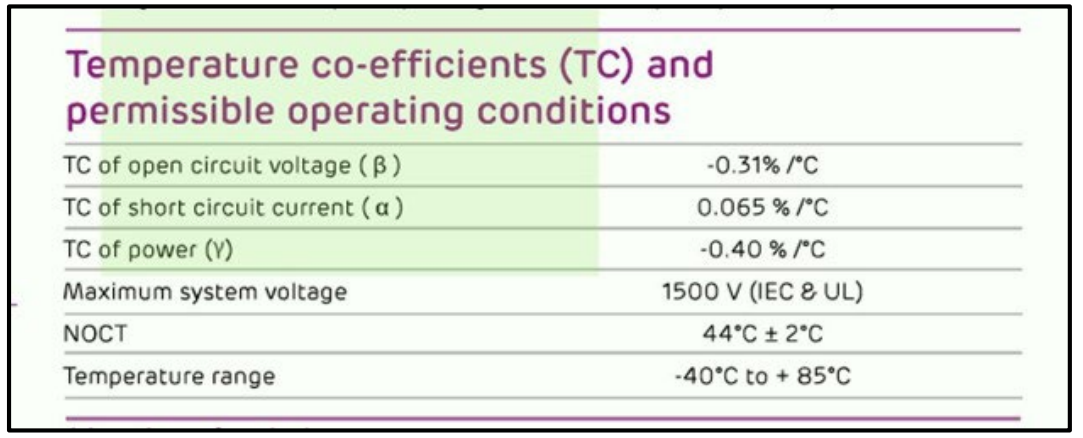
An example of input image



Temperature co-efficients (TC) and permissible operating conditions

TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C

Scaled and cropped image with Gaussian noise

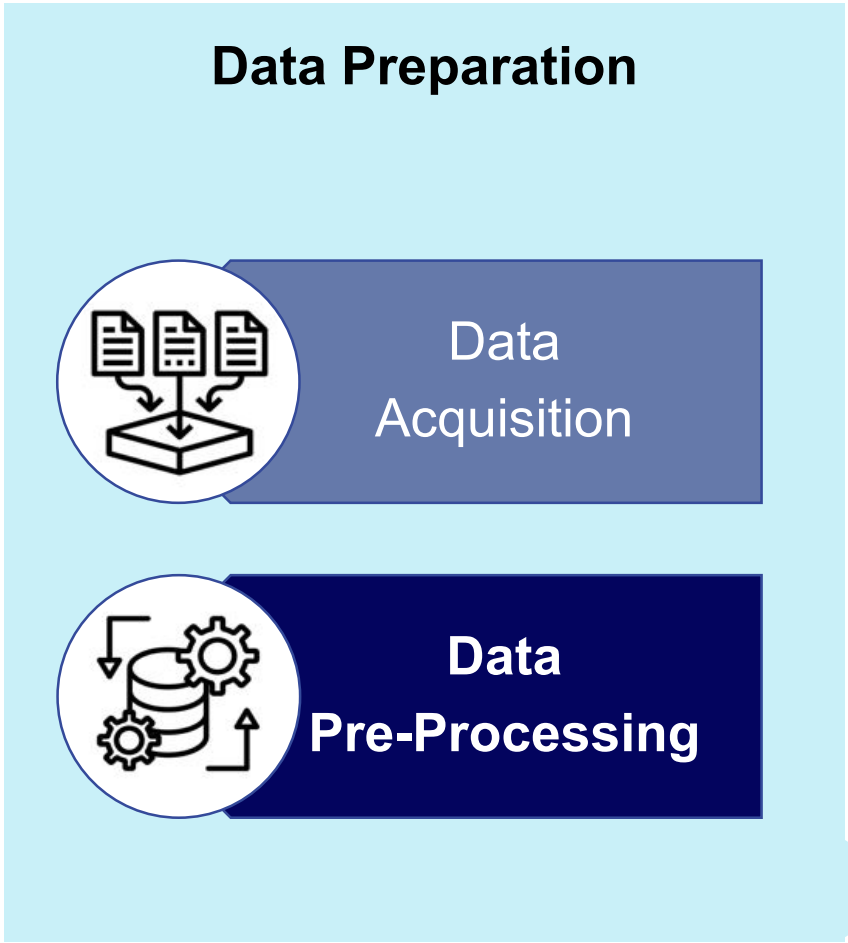


Temperature co-efficients (TC) and permissible operating conditions

TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C
Temperature range	-40°C to + 85°C

Augmented image with Channel shift and random partial mask

Chapter 2 – Approach and Results



2.2. Data Pre-processing

2.2.1. Converting PDF documents to images

- Multi-Page rendering to preserve layout and graphics
- Standardized input

2.2.2. Image Pre-processing

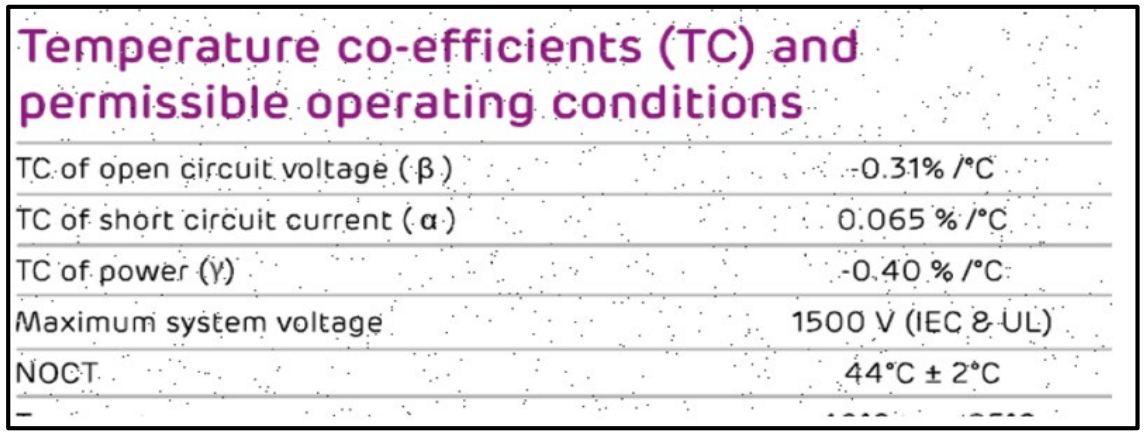
- Enhancement techniques to improve table lines and clarify text

2.2.3. Data extraction using Optical Character Recognition (OCR)

- Extract word tokens and bounding-boxes
- Multi-modal input

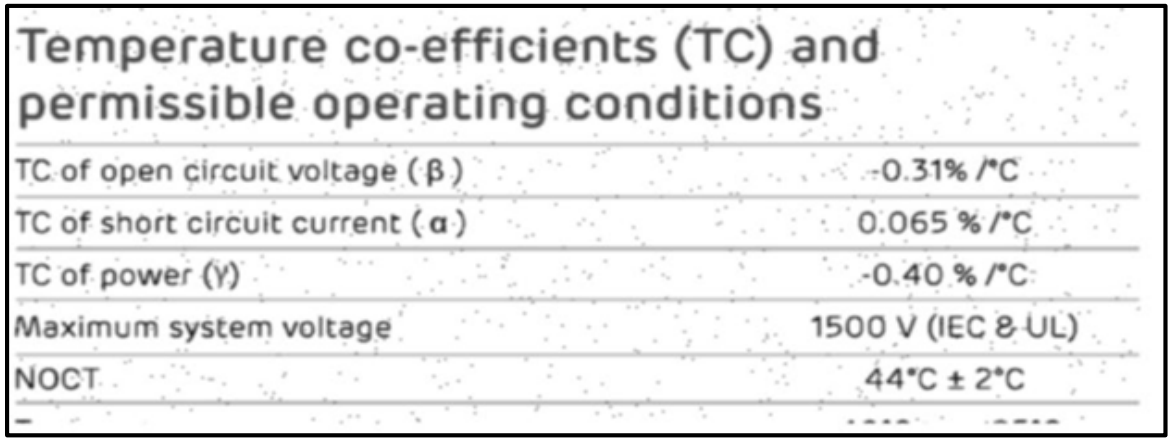
Chapter 2 – Approach and Results

2.2.2. Image Pre-processing



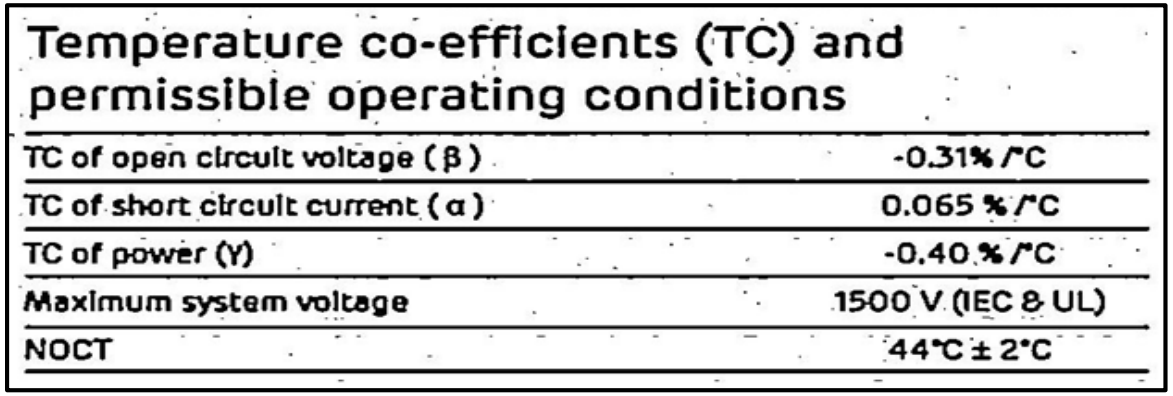
Temperature co-efficients (TC) and permissible operating conditions	
TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C

An example of input image



Temperature co-efficients (TC) and permissible operating conditions	
TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C

Image after applying grayscale conversion and a bilateral filter

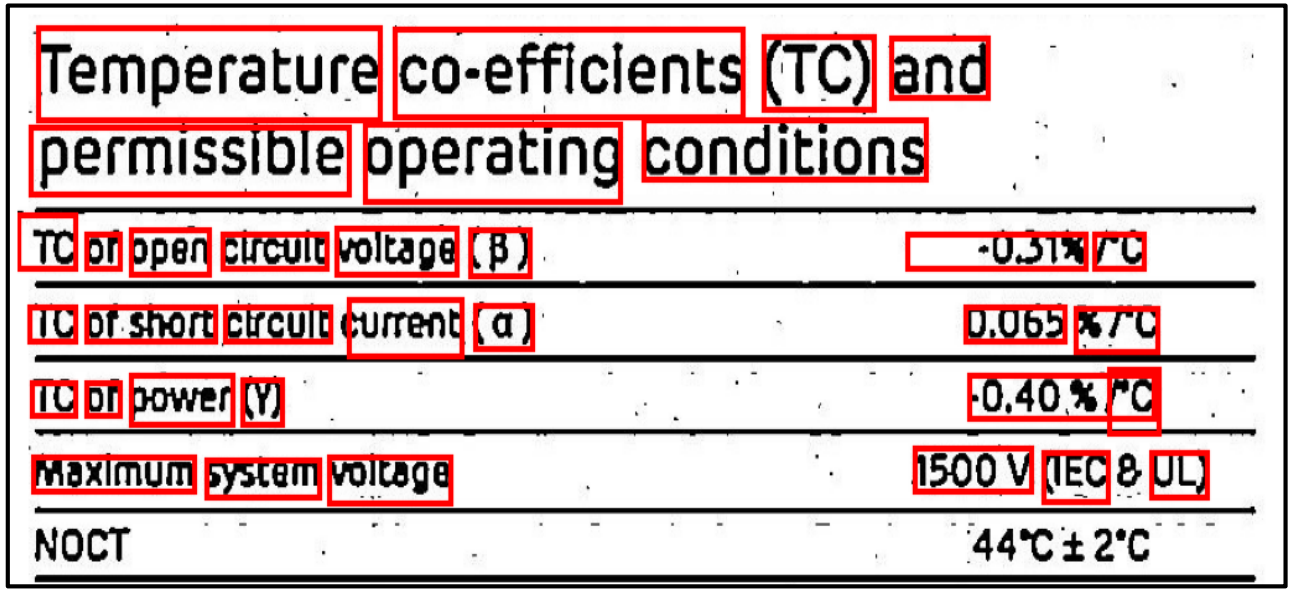


Temperature co-efficients (TC) and permissible operating conditions	
TC of open circuit voltage (β)	-0.31% /°C
TC of short circuit current (α)	0.065 % /°C
TC of power (γ)	-0.40 % /°C
Maximum system voltage	1500 V (IEC & UL)
NOCT	44°C ± 2°C

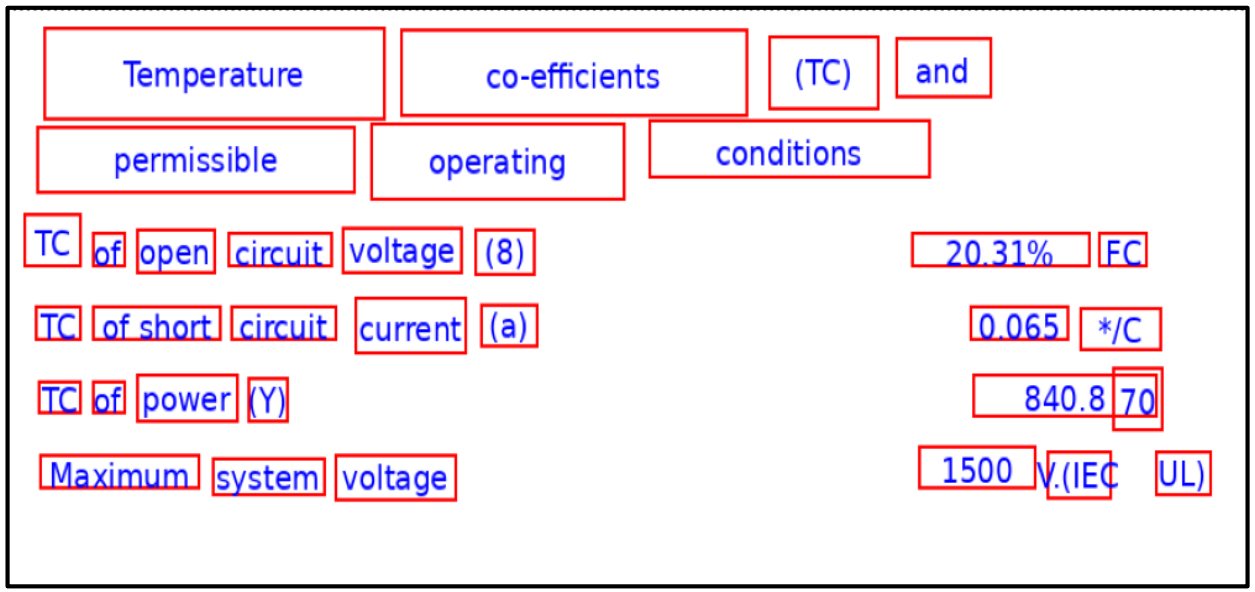
Image after thresholding, morphological opening and sharpening

Chapter 2 – Approach and Results

2.2.3. Data extraction using OCR

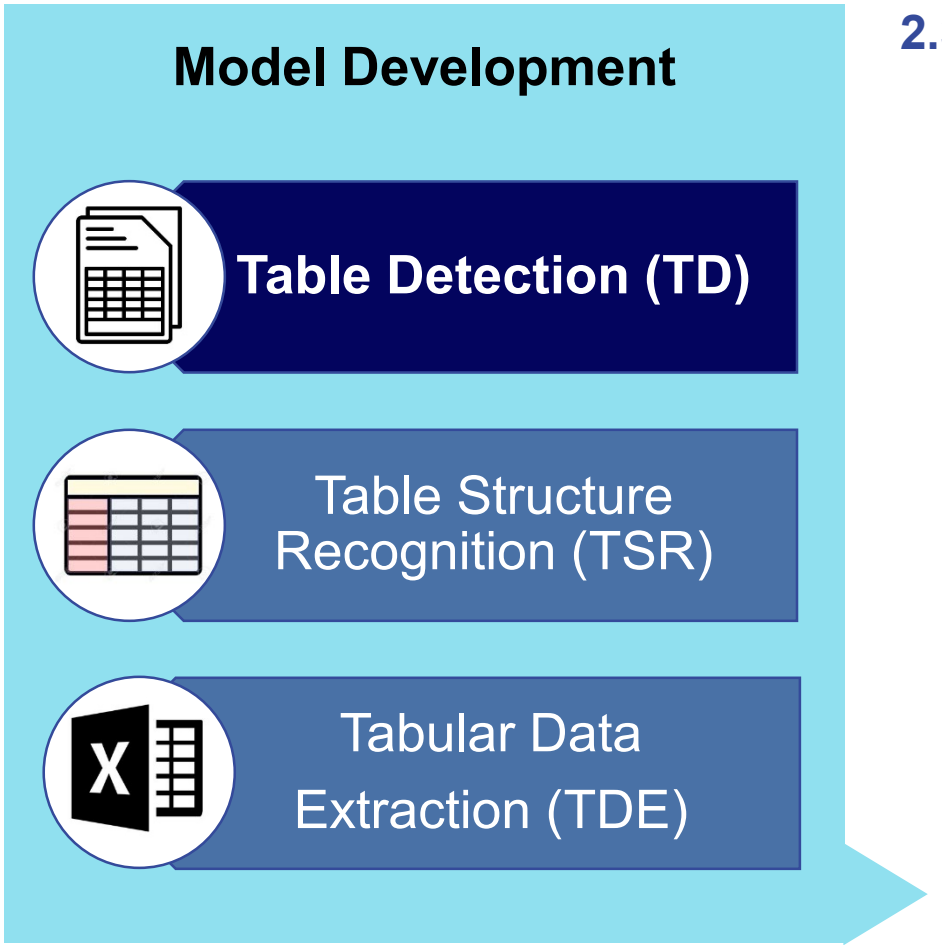


Original image with detected bounding boxes.

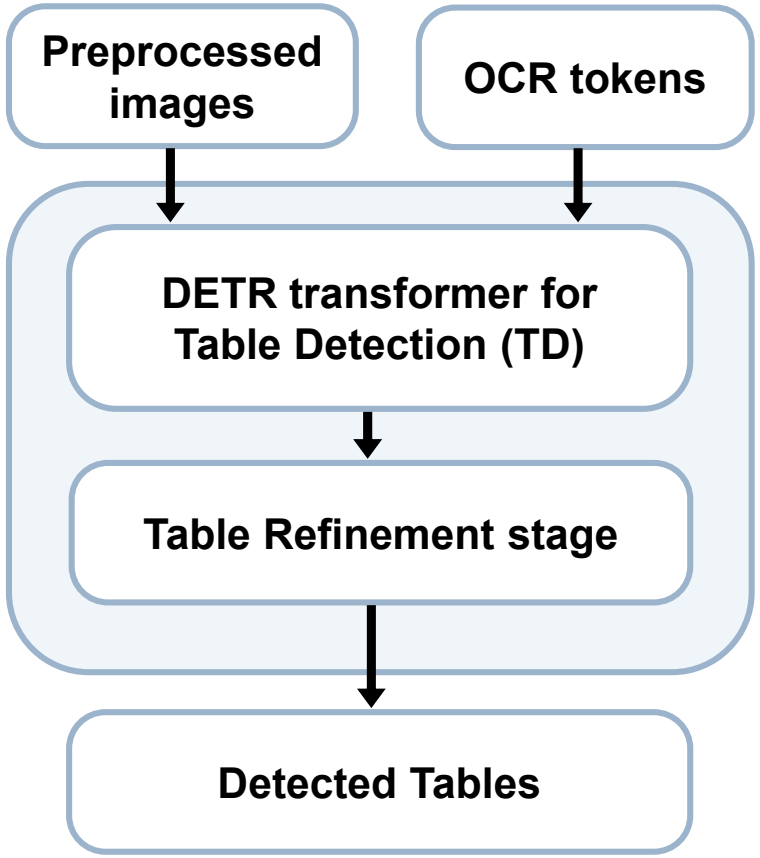


Extracted text tokens represented spatially.

Chapter 2 – Approach and Results

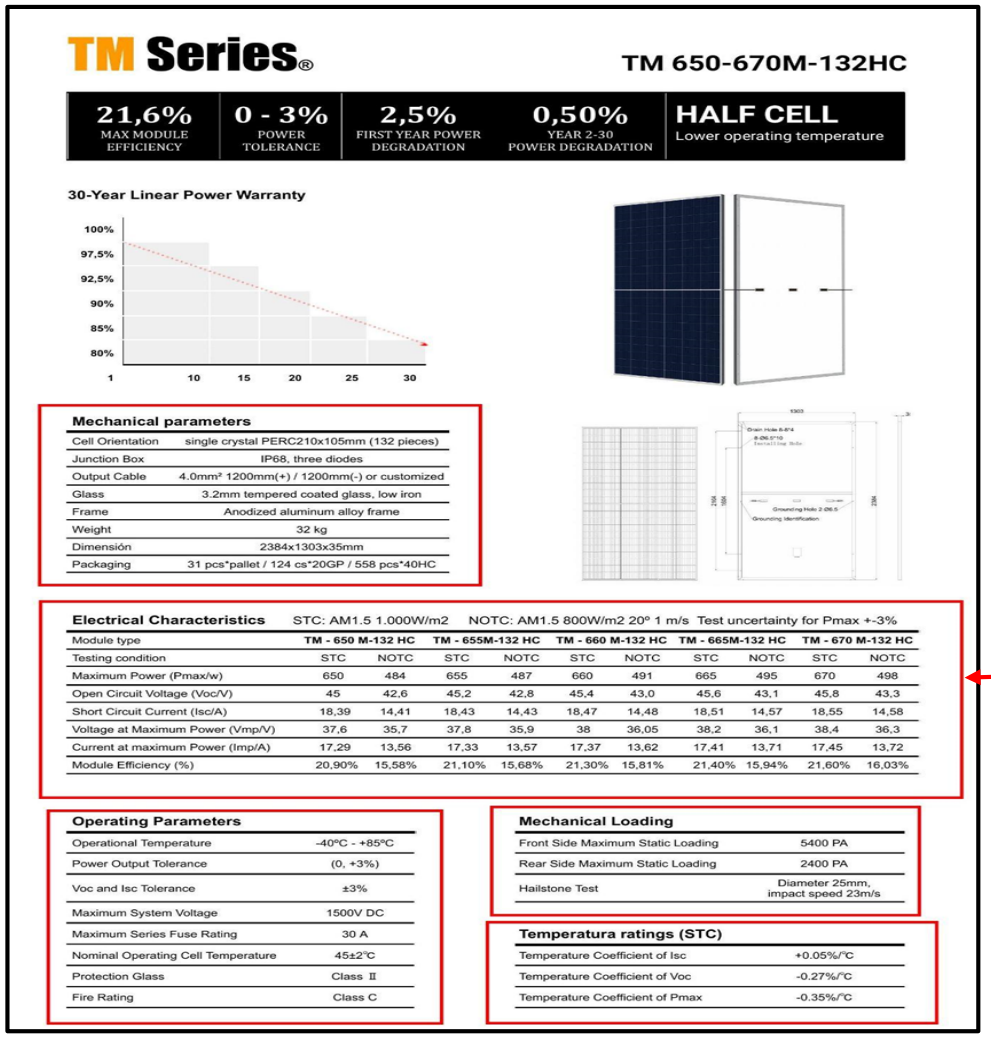


2.3. Table detection (TD)



Chapter 2 – Approach and Results

2.3. Table detection (TD)



TM Series®

TM 650-670M-132HC

21,6%	0 - 3%	2,5%	0,50%	HALF CELL
MAX MODULE EFFICIENCY	POWER TOLERANCE	FIRST YEAR POWER DEGRADATION	YEAR 2-30 POWER DEGRADATION	Lower operating temperature

30-Year Linear Power Warranty

100% 97,5% 92,5% 90% 85% 80%

1 10 15 20 25 30

Mechanical parameters

Cell Orientation	single crystal PERC210x105mm (132 pieces)
Junction Box	IP68, three diodes
Output Cable	4.0mm² 1200mm(+) / 1200mm(-) or customized
Glass	3.2mm tempered coated glass, low iron
Frame	Anodized aluminum alloy frame
Weight	32 kg
Dimensión	2384x1303x35mm
Packaging	31 pcs*pallet / 124 cs*20GP / 558 pcs*40HC

Electrical Characteristics STC: AM1.5 1.000W/m2 NOTC: AM1.5 800W/m2 20º 1 m/s Test uncertainty for Pmax +-3%

Module type	TM - 650 M-132 HC		TM - 655M-132 HC		TM - 660 M-132 HC		TM - 665M-132 HC		TM - 670 M-132 HC	
Testing condition	STC	NOTC	STC	NOTC	STC	NOTC	STC	NOTC	STC	NOTC
Maximum Power (Pmax/w)	650	484	655	487	660	491	665	495	670	498
Open Circuit Voltage (Voc/V)	45	42,6	45,2	42,8	45,4	43,0	45,6	43,1	45,8	43,3
Short Circuit Current (Isc/A)	18,39	14,41	18,43	14,43	18,47	14,48	18,51	14,57	18,55	14,58
Voltage at Maximum Power (Vmp/V)	37,6	35,7	37,8	35,9	38	36,05	38,2	36,1	38,4	36,3
Current at maximum Power (Imp/A)	17,29	13,56	17,33	13,57	17,37	13,62	17,41	13,71	17,45	13,72
Module Efficiency (%)	20,90%	15,58%	21,10%	15,68%	21,30%	15,81%	21,40%	15,94%	21,60%	16,03%

Operating Parameters

Operational Temperature	-40ºC - +85ºC
Power Output Tolerance	(0, +3%)
Voc and Isc Tolerance	±3%
Maximum System Voltage	1500V DC
Maximum Series Fuse Rating	30 A
Nominal Operating Cell Temperature	45±2ºC
Protection Glass	Class Ⅱ
Fire Rating	Class C

Mechanical Loading

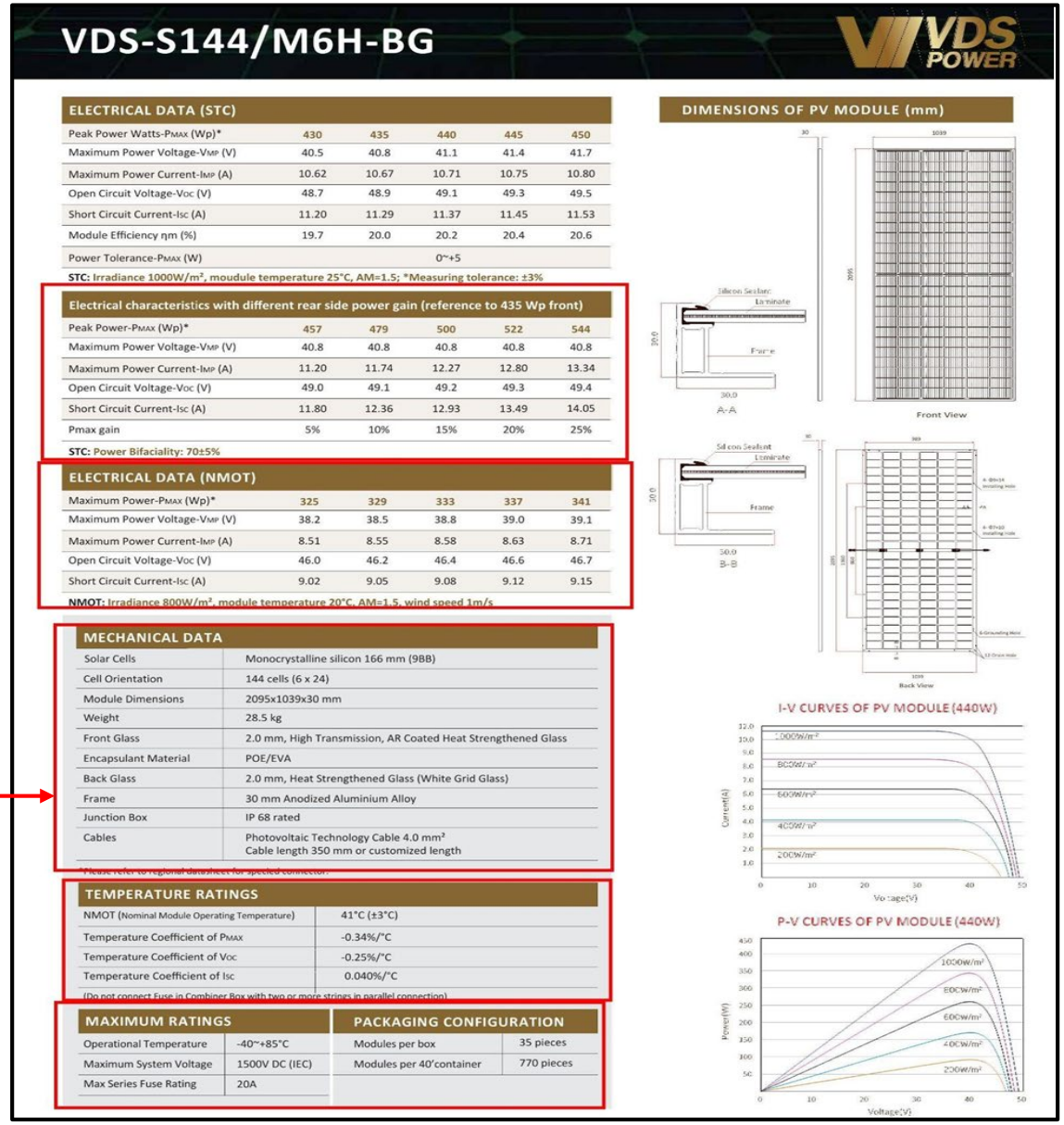
Front Side Maximum Static Loading	5400 PA
Rear Side Maximum Static Loading	2400 PA
Hailstone Test	Diameter 25mm, impact speed 23m/s

Temperatura ratings (STC)

Temperature Coefficient of Isc	+0.05%/ºC
Temperature Coefficient of Voc	-0.27%/ºC
Temperature Coefficient of Pmax	-0.35%/ºC

AP75 score – 64.2%

Red Bounding boxes indicate the tables detected by the TD model.



VDS-S144/M6H-BG

VDS POWER

ELECTRICAL DATA (STC)

Peak Power Watts-Pmax (Wp)*	430	435	440	445	450
Maximum Power Voltage-Vmp (V)	40.5	40.8	41.1	41.4	41.7
Maximum Power Current-Imp (A)	10.62	10.67	10.71	10.75	10.80
Open Circuit Voltage-Voc (V)	48.7	48.9	49.1	49.3	49.5
Short Circuit Current-Isc (A)	11.20	11.29	11.37	11.45	11.53
Module Efficiency ηm (%)	19.7	20.0	20.2	20.4	20.6
Power Tolerance-Pmax (W)	0~+5				

STC: Irradiance 1000W/m², moudule temperature 25°C, AM=1.5; *Measuring tolerance: ±3%

Electrical characteristics with different rear side power gain (reference to 435 Wp front)

Peak Power-Pmax (Wp)*	457	479	500	522	544
Maximum Power Voltage-Vmp (V)	40.8	40.8	40.8	40.8	40.8
Maximum Power Current-Imp (A)	11.20	11.74	12.27	12.80	13.34
Open Circuit Voltage-Voc (V)	49.0	49.1	49.2	49.3	49.4
Short Circuit Current-Isc (A)	11.80	12.36	12.93	13.49	14.05
Pmax gain	5%	10%	15%	20%	25%

STC: Power Bifaciality: 70±5%

ELECTRICAL DATA (NMOT)

Maximum Power-Pmax (Wp)*	325	329	333	337	341
Maximum Power Voltage-Vmp (V)	38.2	38.5	38.8	39.0	39.1
Maximum Power Current-Imp (A)	8.51	8.55	8.58	8.63	8.71
Open Circuit Voltage-Voc (V)	46.0	46.2	46.4	46.6	46.7
Short Circuit Current-Isc (A)	9.02	9.05	9.08	9.12	9.15

NMOT: Irradiance 800W/m², module temperature 20°C, AM=1.5, wind speed 1m/s

MECHANICAL DATA

Solar Cells	Monocrystalline silicon 166 mm (9BB)
Cell Orientation	144 cells (6 x 24)
Module Dimensions	2095x1039x30 mm
Weight	28.5 kg
Front Glass	2.0 mm, High Transmission, AR Coated Heat Strengthened Glass
Encapsulant Material	POE/EVA
Back Glass	2.0 mm, Heat Strengthened Glass (White Grid Glass)
Frame	30 mm Anodized Aluminium Alloy
Junction Box	IP 68 rated
Cables	Photovoltaic Technology Cable 4.0 mm² Cable length 350 mm or customized length

TEMPERATURE RATINGS

NMOT (Nominal Module Operating Temperature)	41°C (±3°C)
Temperature Coefficient of Pmax	-0.34%/°C
Temperature Coefficient of Voc	-0.25%/°C
Temperature Coefficient of Isc	0.040%/°C

(Do not connect Fuse in Combiner Box with two or more strings in parallel connection)

MAXIMUM RATINGS

Operational Temperature	-40~+85°C
Maximum System Voltage	1500V DC (IEC)
Max Series Fuse Rating	20A

PACKAGING CONFIGURATION

Modules per box	35 pieces
Modules per 40'container	770 pieces

DIMENSIONS OF PV MODULE (mm)

Front View

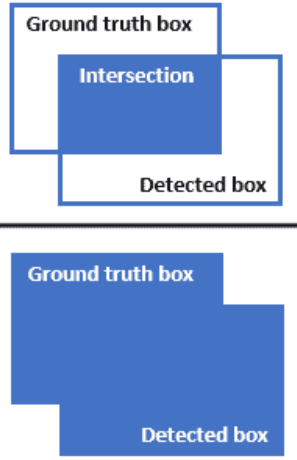
I-V CURVES OF PV MODULE (440W)

P-V CURVES OF PV MODULE (440W)

AP75 score – 38.3%

Chapter 2 – Approach and Results

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} = \frac{\text{Intersection}}{\text{Ground truth box} + \text{Detected box}}$$



2.3. Table Detection (TD)

- PubTables-1M dataset - 460,589 annotated document pages
- PV module training dataset – 169 images Test dataset – 15 images

	AP50	AP75	AP	AR
Trained on PubTables-1M dataset	50.7%	14.1%	24.9%	36.8%
Fine-tuned on PV module training dataset	87.8%	80.5%	66.5%	75.4%
Fine-tuned on PV module training dataset including OCR tokens	87.1%	81.8%	67.3%	74.3%
Fine-tuned on modified PV module training dataset with negative samples	88.7%	81.1%	66.0%	72.6%
Performing Hyperparameter optimization	89.5%	80.9%	68.6%	77.5%

Chapter 2 – Approach and Results

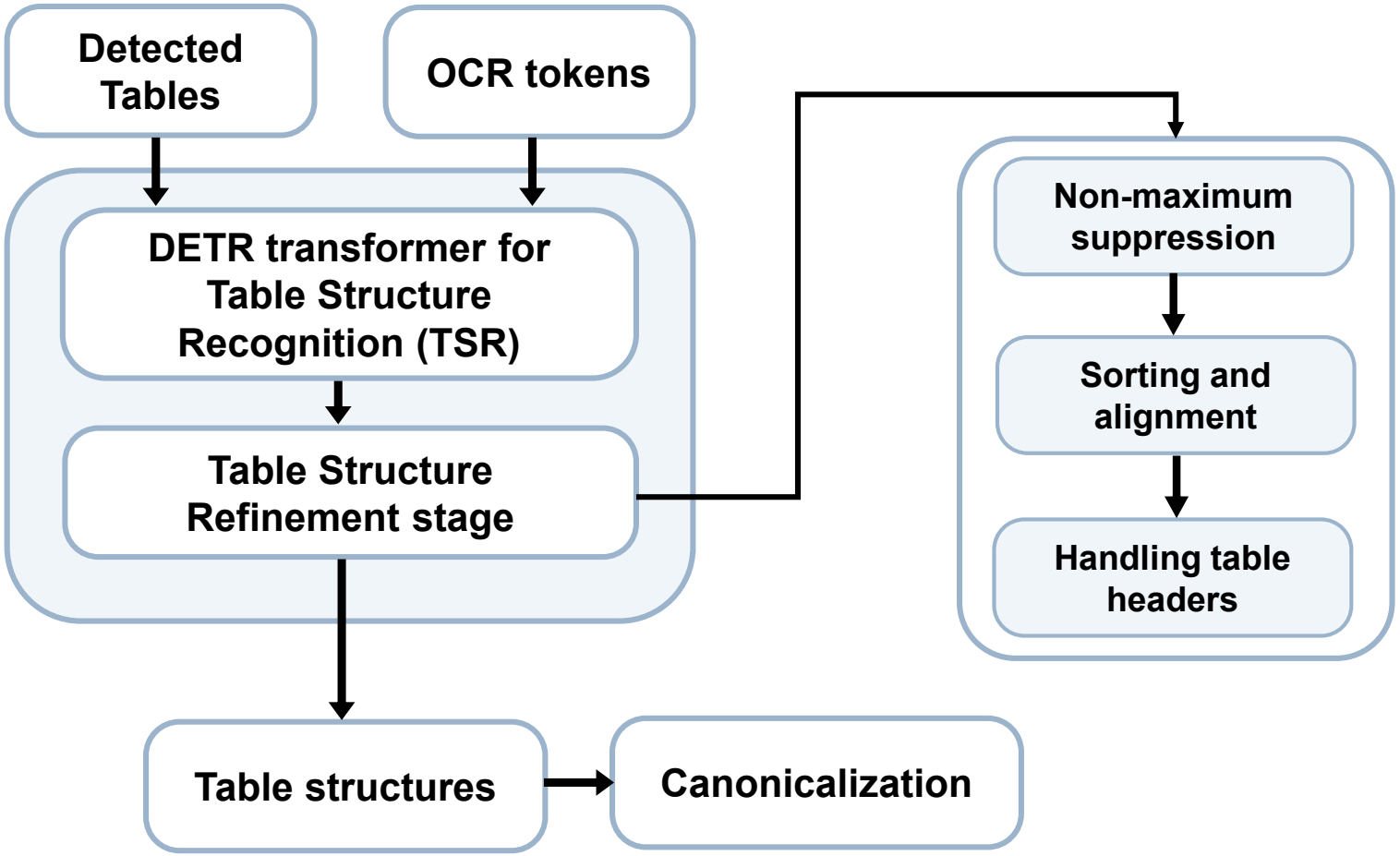
Model Development

Table Detection (TD)

Table Structure Recognition (TSR)

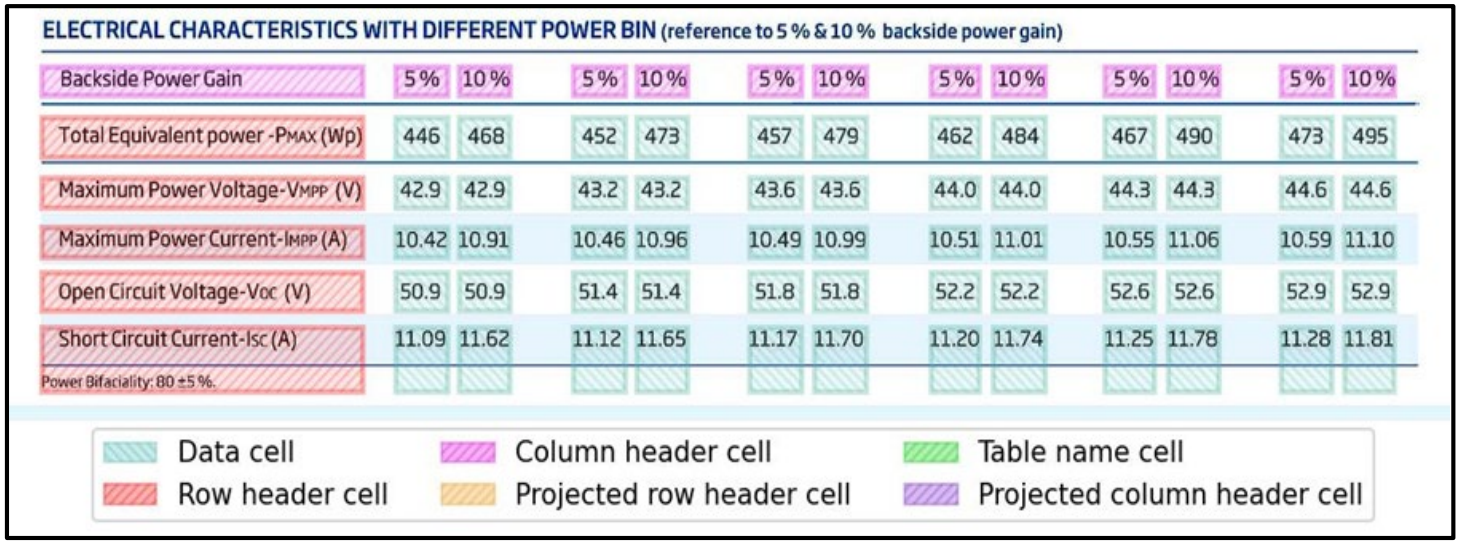
Tabular Data Extraction (TDE)

2.4. Table Structure Recognition (TSR)



Chapter 2 - Approach and Results

2.4. Table Structure Recognition (TSR)



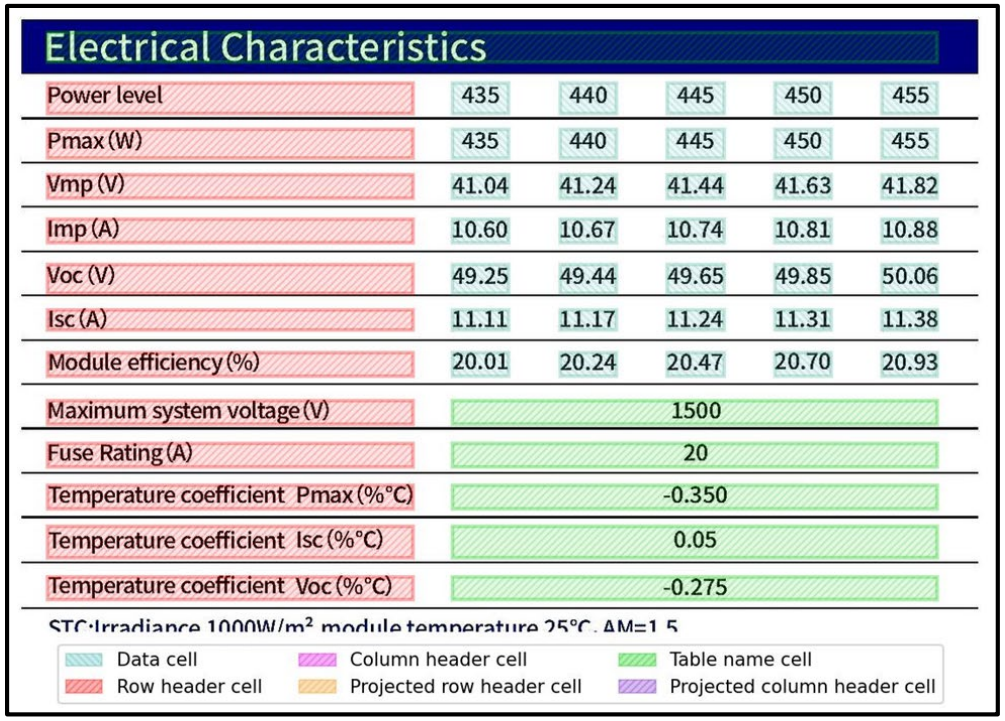
ELECTRICAL CHARACTERISTICS WITH DIFFERENT POWER BIN (reference to 5 % & 10 % backside power gain)

Backside Power Gain	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%	5%	10%
Total Equivalent power -Pmax (Wp)	446	468	452	473	457	479	462	484	467	490	473	495
Maximum Power Voltage-Vmpp (V)	42.9	42.9	43.2	43.2	43.6	43.6	44.0	44.0	44.3	44.3	44.6	44.6
Maximum Power Current-Impp (A)	10.42	10.91	10.46	10.96	10.49	10.99	10.51	11.01	10.55	11.06	10.59	11.10
Open Circuit Voltage-Voc (V)	50.9	50.9	51.4	51.4	51.8	51.8	52.2	52.2	52.6	52.6	52.9	52.9
Short Circuit Current-Isc (A)	11.09	11.62	11.12	11.65	11.17	11.70	11.20	11.74	11.25	11.78	11.28	11.81

Power Bifaciality: 80 ±5 %.

Data cell | Column header cell | Table name cell
Row header cell | Projected row header cell | Projected column header cell

AP75 score – 47.1%



Electrical Characteristics

Power level	435	440	445	450	455
Pmax (W)	435	440	445	450	455
Vmp (V)	41.04	41.24	41.44	41.63	41.82
Imp (A)	10.60	10.67	10.74	10.81	10.88
Voc (V)	49.25	49.44	49.65	49.85	50.06
Isc (A)	11.11	11.17	11.24	11.31	11.38
Module efficiency (%)	20.01	20.24	20.47	20.70	20.93
Maximum system voltage (V)	1500				
Fuse Rating (A)	20				
Temperature coefficient Pmax (%°C)	-0.350				
Temperature coefficient Isc (%°C)	0.05				
Temperature coefficient Voc (%°C)	-0.275				

STC:Irradiance 1000W/m² module temperature 25°C, AM=1.5

Data cell | Column header cell | Table name cell
Row header cell | Projected row header cell | Projected column header cell

AP75 score – 71.0%

Chapter 2 – Approach and Results

2.4. Table Structure Recognition (TSR)

- PubTables-1M dataset - 947,642 fully annotated tables
- PV module training dataset – 545 table images Test dataset – 65 table images

	AP50	AP75	AP	AR
Trained on PubTables-1M dataset	10.4%	6.8%	6.2%	6.7%
Fine-tuned on PV module training dataset	9.1%	8.4%	6.9%	7.5%
Fine-tuned on PV module training dataset with enhances canonicalization algorithm	61.8%	50.3%	43.1%	56.9%
Performing Hyperparameter optimization	71.3%	57.3%	49.0%	58.3%

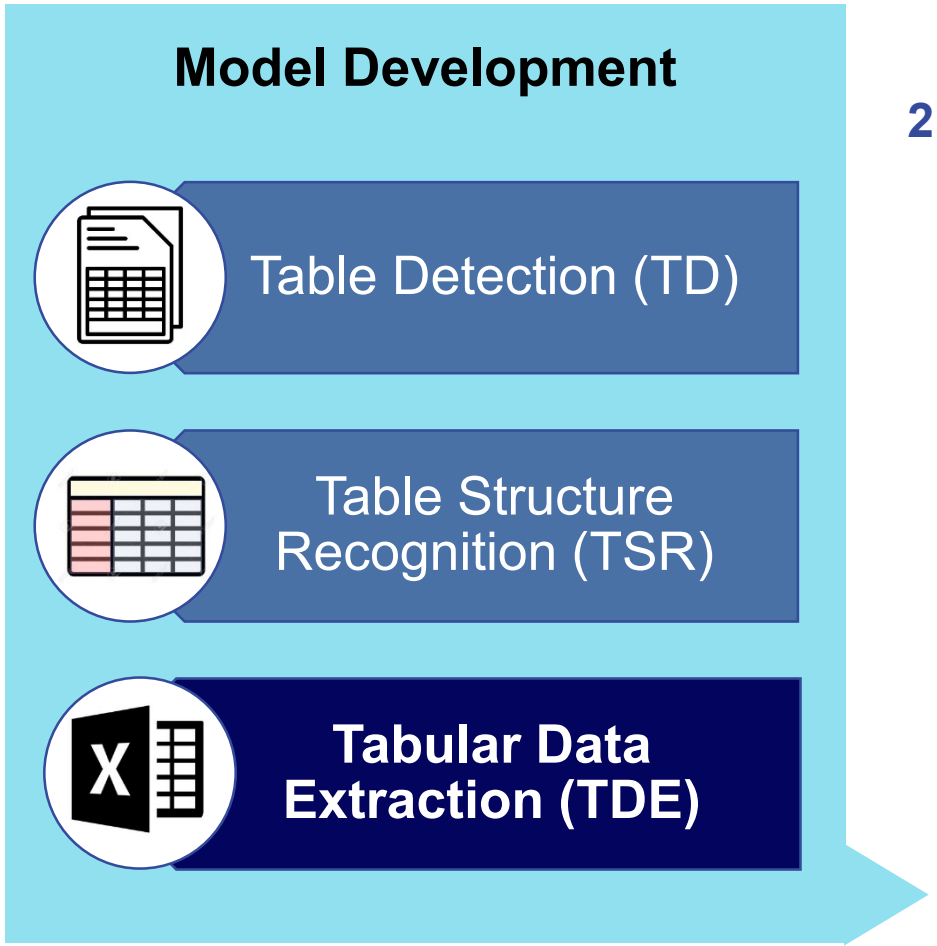
Chapter 2 - Approach and Results

2.4. Evaluation Metric - Grid Table Similarity (GriTS)

- Assess model's accuracy in capturing table layout, content location, and cell relationships
- Variants - Topological GriTS , Content GriTS & Location GriTS

	Table Type	Acc_con	GriTS_Top	GriTS_Con	GriTS_Loc
Fine-tuned model with Hyperparameter Optimizaion	Simple tables	83.21%	89.18%	83.21%	77.13%
	Complex tables	95.03%	94.93%	95.03%	92.18%
	All tables	91.89%	93.40%	91.89%	88.18%

Chapter 2 – Approach and Results



2.5. Tabular Data Extraction (TDE)

- Simplify table structures assuming - row/column alignment and logically consistent tables
- Duplicate cell content across the corresponding rows/columns that a merged cell spans
- Extract and export data to CSV/Excel files for further analysis

Chapter 2 - Approach and Results

Table structure recognized by the TSR model

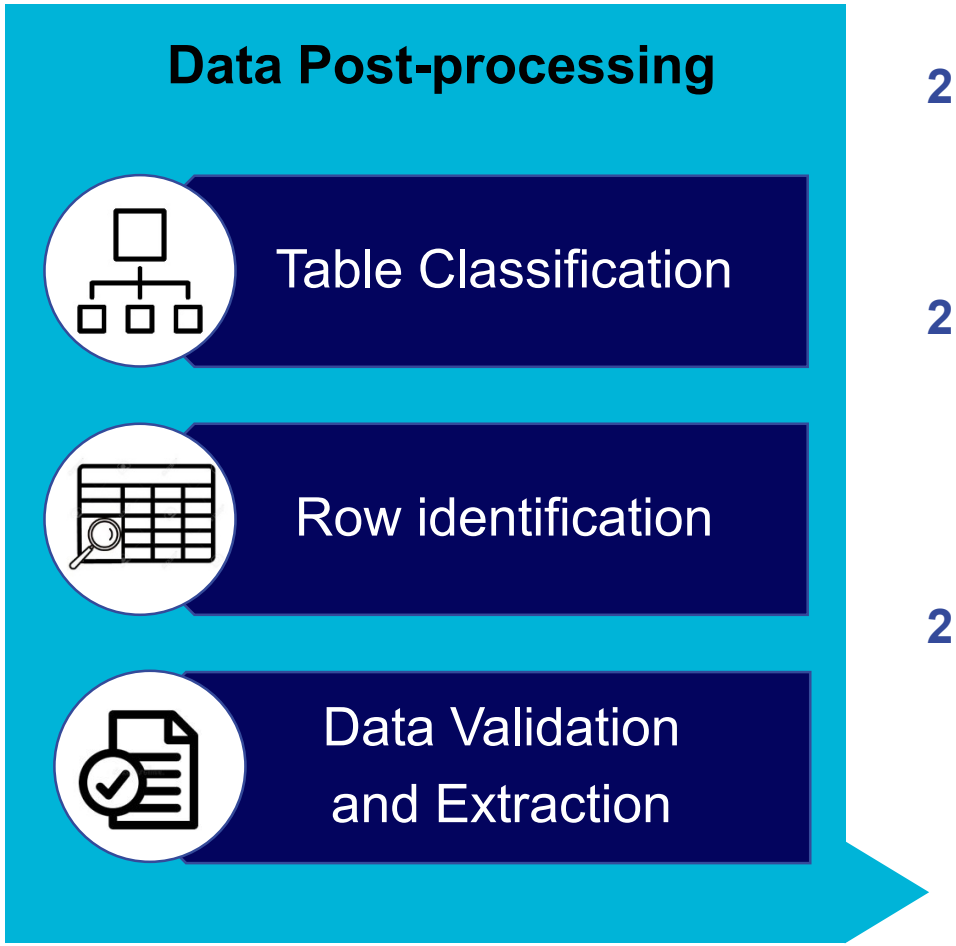
Electrical Parameters

Module Type	SPICN6(LAR)-60-375/IH		SPICN6(LAR)-60-380/IH		SPICN6(LAR)-60-385/IH	
	STC	NOCT	STC	NOCT	STC	NOCT
Maximum Power (Pmax/W)	375	282	380	286	385	290
Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40.0
Short Circuit Current (Isc/A)	11.50	9.23	11.61	9.32	11.72	9.42
Module Efficiency	20.9%		21.2%		21.5%	

Extracted data and exported to Excel

	A	B	C	D	E	F	G
1	Module Type 20.0kg	SPICN6(LAR)}-60-375/IH | SIE	SPICN6(LAR)}-60-375/IH | NOCT	SPICN6(LAR)-60-380/IH | STC	SPICN6(LAR)-60-380/IH | NOCT	SPICN6(LAR)-60-385/IH | STC	SPICN6(LAR)-60-385/IH | NOCT
2							
3	Maximum Power (Pmax/W)	375	282	380	286	385	290
4							
5	Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
6							
7	Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
8							
9	Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40
10							
11	Short Circuit Current (Isc/A)	11.5	9.23	11.61	9.32	11.72	9.42
12							
13	Module Efficiency	20.90%	20.90%	21.20%	21.20%	21.50%	21.50%

Chapter 2 – Approach and Results



2.6. Table classification

- Multinomial Naïve Bayes classifier with a TF-IDF word vectorizer.

2.7. Row / Column identification

- Regular Expression Pattern Matching
- Large Language Models

2.7. Data Validation and Extraction

- Regular expressions to validate and extract relevant data
- Output saved in Excel format

Chapter 2 – Approach and Results

2.6. Table classification

- Overall Accuracy - **87%**
- Strong performance distinguishing “Thermal,” “Mechanical,” and “Packaging” tables
- Poor classification performance on Electrical characteristics table at “STC” vs. “NMOT” conditions due to nearly identical keywords

Class	Precision	Recall	F1-Score
Electrical Characteristics at Standard Testing Conditions (STC)	0.75	0.60	0.67
Electrical Characteristics at Nominal Module Operating Temperature (NMOT)	0.33	0.67	0.44
Thermal Characteristics	1.00	0.88	0.93
Mechanical Characteristics	1.00	1.00	1.00
Packaging	0.80	1.00	0.89
Others	1.00	0.94	0.97

Chapter 2 – Approach and Results

2.7. Row / Column identification

Electrical Characteristics	STC: AM1.5 1.000W/m2		NOTC: AM1.5 800W/m2 20° 1 m/s Test uncertainty for Pmax +-3%							
Module type	TM - 650 M-132 HC		TM - 655M-132 HC		TM - 660 M-132 HC		TM - 665M-132 HC		TM - 670 M-132 HC	
Testing condition	STC	NOTC	STC	NOTC	STC	NOTC	STC	NOTC	STC	NOTC
Maximum Power (Pmax/w)	650	484	655	487	660	491	665	495	670	498
Open Circuit Voltage (Voc/V)	45	42,6	45,2	42,8	45,4	43,0	45,6	43,1	45,8	43,3
Short Circuit Current (Isc/A)	18,39	14,41	18,43	14,43	18,47	14,48	18,51	14,57	18,55	14,58
Voltage at Maximum Power (Vmp/V)	37,6	35,7	37,8	35,9	38	36,05	38,2	36,1	38,4	36,3
Current at maximum Power (Imp/A)	17,29	13,56	17,33	13,57	17,37	13,62	17,41	13,71	17,45	13,72
Module Efficiency (%)	20,90%	15,58%	21,10%	15,68%	21,30%	15,81%	21,40%	15,94%	21,60%	16,03%

Regular expression – ‘eff((?!code).)*$|ncell|model\(%\)’
Efficiency data extracted using Regular expression – ["20.90%", "15.58%", "21.10%", "15.68%", "21.30%", "15.81%", "21.40%", "15.94%", "21.60%", "16.03%“]

Natural Language Prompt – ‘From the electrical characteristics table, extract the module efficiency data (in percentage) for Standard Testing Conditions (STC) only. Round the value to two decimal places.’
Efficiency extracting using LLM – ["20.90%", "21.10%", "21.30%", "21.40%“, "21.60%"]

Chapter 2 – Approach and Results

2.7. Data Validation and Extraction

Data extraction accuracy of the complete pipeline on test set consisting of 10 PDF files – 52.66% AP

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		name	year	length	width	E/eff	E/pmpp	E/vmpp	E/impp	E/voc	E/isc	EN/pmpp	EN/vmpp	EN/impp	EN/voc	EN/isc	T/isc	T/pmpp	T/voc	/pcs_palle	/pallet_con
2	0	5-VDS	2023	2095	1039	19.7	430	40.5	10.62	48.7	11.20	325.0	38.2	8.51	46.0	9.02	0.040	-0.34%/°	-0.25%/°	35	770
3	1	5-VDS	2023	2095	1039	20.0	435	40.8	10.67	48.9	11.29	329.0	38.5	8.55	46.2	9.03	0.040	-0.34%/°	-0.25%/°	35	770
4	2	5-VDS	2023	2095	1039	20.2	440	41.1	10.71	49.1	11.37	333.0	38.8	8.58	46.4	9.08	0.040	-0.34%/°	-0.25%/°	35	770
5	3	5-VDS	2023	2095	1039	20.4	445	41.4	10.75	49.3	11.45	337.0	39.0	8.63	46.6	9.12	0.040	-0.34%/°	-0.25%/°	35	770
6	4	5-VDS	2023	2095	1039	20.6	450	41.7	10.80	49.5	11.53	341.0	39.1	8.71	46.7	9.15	0.040	-0.34%/°	-0.25%/°	35	770

Example of Manually extracted data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1		name	year	length	width	E/eff	E/pmpp	E/vmpp	E/impp	E/voc	E/isc	EN/pmpp	EN/vmpp	EN/impp	EN/voc	EN/isc	T/isc	T/pmpp	T/voc	/pcs_palle	pallet_con
2	0	5-VDS	2023	2095	1039		457	40.8	11.20	49.0		325.0	38.2	8.51	46.0	9.02	0.040%/°	-0.34%/°	-0.25%/°		770
3	1	5-VDS	2023	2095	1039		479	40.8	11.74	49.1		329.0	38.5	8.55	46.2	9.05	0.040%/°	-0.34%/°	-0.25%/°		770
4	2	5-VDS	2023	2095	1039		500	40.8	12.27	49.2		333.0	38.8	8.58	46.4	9.08	0.040%/°	-0.34%/°	-0.25%/°		770
5	3	5-VDS	2023	2095	1039		522	40.8	12.80	49.3		337.0	39.0	8.63	46.6	9.12	0.040%/°	-0.34%/°	-0.25%/°		770
6	4	5-VDS	2023	2095	1039		544	40.8	13.34	49.4		341.0	39.1	8.71	46.7	9.15	0.040%/°	-0.34%/°	-0.25%/°		770

Example of data extracted using the pipeline

Chapter 3
Summary & Outlook

Chapter 3 - Summary and Outlook

3.1. Summary

- Integrated DETR with OCR tokens to enable multi-model input for table TD and TSR tasks
- Enhanced canonicalization algorithm for handling horizontal and dual-axis tables
- Robust handling of merged cells, multi-row/multi-column headers, and image-based PDFs
- Streamlined data extraction pipeline and improved its reliability
- Quicker, large-scale analysis of PV module datasheets for improved decision-making
- Demonstrated feasibility and strong potential for broader application in other domains with complex tables

Chapter 3 - Summary and Outlook

3.2. Limitations

- Dependence on OCR extraction quality
- Capturing complex table structures in densely-packed tables
- Misclassifying tables with similar headers or minimal textual cues

3.3. Future work

- Advanced OCR engines or domain-focused fine-tuning OCR
- Advanced models to perform context-aware validation and extraction of data
- Expanding dataset coverage to handle even more diverse table layouts

universität freiburg

Thank you for your attention!

universität freiburg

1st Examiner :
Prof. Dr. Hannah Bast
University of Freiburg, Freiburg im Breisgau

2nd Examiner :
Prof. Dr. Holger Neuhaus

Advisers :
Dr. Ing. Christian Reichel

Dr. Patrick Brosi
University of Freiburg, Freiburg im Breisgau

Questions

Appendix

Chapter 1 - Introduction

1.0. Importance of automated data extraction

- Widespread Industrial Applications – Healthcare, legal, Financial sectors
- Scalable, efficient, and saves operational costs
- Enables Informed decisions

Information Extraction

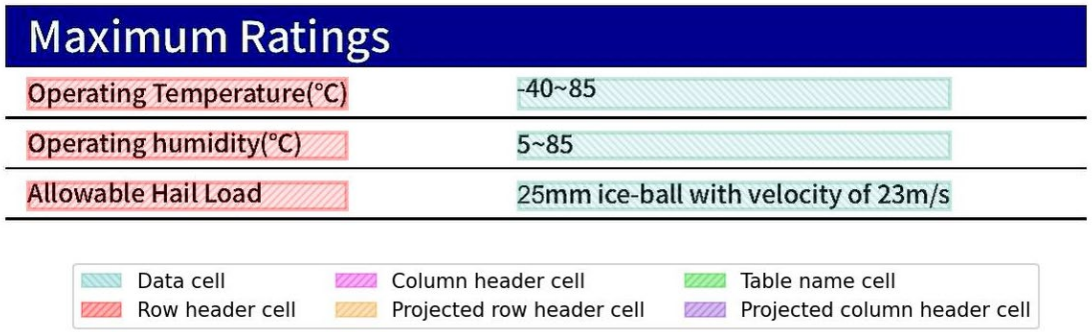
Chapter 2 – Approach and Results

2.1.1. Manual annotation

Maximum Ratings	
Operating Temperature(°C)	-40~85
Operating humidity(°C)	5~85
Allowable Hail Load	25mm ice-ball with velocity of 23m/s

Example of a Manually annotated row

Maximum Ratings	
Operating Temperature(°C)	-40~85
Operating humidity(°C)	5~85
Allowable Hail Load	25mm ice-ball with velocity of 23m/s



An example of a table with structures recognized by the TSR model

Maximum Ratings	
Operating Temperature(°C)	-40~85
Operating humidity(°C)	5~85
Allowable Hail Load	25mm ice-ball with velocity of 23m/s

Example of a Manually annotated column

Maximum Ratings	
Operating Temperature(°C)	-40~85
Operating humidity(°C)	5~85
Allowable Hail Load	25mm ice-ball with velocity of 23m/s

The corresponding cell derived

Chapter 2 – Approach and Results

2.1.1. Manual annotation

.

Electrical Properties (NMOT)

Model		LG350N1C-V5
Maximum Power (Pmax)	[W]	261
MPP Voltage (Vmpp)	[V]	33.0
MPP Current (Impp)	[A]	7.91
Open Circuit Voltage (Voc)	[V]	38.80
Short Circuit Current (Isc)	[A]	8.53

Data cell | Column header cell | Table name cell
Row header cell | Projected row header cell | Projected column header cell

Since Predicted bounding box is fully contained in the ground truth, IoU<1

Electrical Properties (NMOT)

Model		LG350N1C-V5
Maximum Power (Pmax)	[W]	261
MPP Voltage (Vmpp)	[V]	33.0
MPP Current (Impp)	[A]	7.91
Open Circuit Voltage (Voc)	[V]	38.80
Short Circuit Current (Isc)	[A]	8.53

Example of a Manually annotated row

Electrical Properties (NMOT)

Model		LG350N1C-V5
Maximum Power (Pmax)	[W]	261
MPP Voltage (Vmpp)	[V]	33.0
MPP Current (Impp)	[A]	7.91
Open Circuit Voltage (Voc)	[V]	38.80
Short Circuit Current (Isc)	[A]	8.53

Example of a Manually annotated column

Electrical Properties (NMOT)

Model		LG350N1C-V5
Maximum Power (Pmax)	[W]	261
MPP Voltage (Vmpp)	[V]	33.0
MPP Current (Impp)	[A]	7.91
Open Circuit Voltage (Voc)	[V]	38.80
Short Circuit Current (Isc)	[A]	8.53

The corresponding cell derived

Chapter 2 - Approach and Results

2.1.2. Image augmentation

Type	US280M60B	US290M60B	US300M60B
Pmax [W]	280	290	300
Vmpp [V]	32.94	33.32	33.72
Impp [A]	8.50	8.72	8.93
Voc [V]	39.06	39.46	39.87
Isc [A]	9.07	9.18	9.29
Max. System Voltage	1000V DC	1000V DC	1000V DC

Data cell Column header cell Projected row header cell

Table structures initially recognized without augmentation

Chapter 2 - Approach and Results

2.2.2 Image Preprocessing and OCR enhancement

Electrical Parameters

Module Type	SPICN6(LAR)-60-375/IH		SPICN6(LAR)-60-380/IH		SPICN6(LAR)-60-385/IH	
	STC	NOCT	STC	NOCT	STC	NOCT
Maximum Power (Pmax/W)	375	282	380	286	385	290
Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40.0
Short Circuit Current (Isc/A)	11.50	9.23	11.61	9.32	11.72	9.42
Module Efficiency	20.9%		21.2%		21.5%	

Table with its internal structure recognized

	A	B	C	D	E	F	G
1	Module Type 20.0kg	SPICN6(LAR)}-60-375/IH | SIC	SPICN6(LAR)}-60-375/IH | NOCT	SPICN6(LAR)-60-380/1H | STC	SPICN6(LAR)-60-380/IH | NOCT	SPICN6(LAR)-60-385/IH | S1E	SPICN6(LAR)-60-385/IH | NOCT
2							
3	Maximum Power (Pmax/W)	375	282	380	286	385	290
4							
5	Maximum Power Voltage (Vmpp/V)	Ee|	525	35:03:00	32.7	35:05:00	329
6							
7	Maximum Power Current (Impp/A)	10.69	8.68	10.77	875	10.85	8.82
8							
9	Open Circuit Voltage (Voc/V)	41.6	39.8	ALT	39.9	A18	40
10							
11	Short Circuit Current (Isc/A)	11.5	9.23	11.61	9.32	11.72	9.42
12							
13	Module Efficiency	20.90%	20.90%	21.20%	21.20%	21.50%	21.50%

Data extracted before image preprocessing

	A	B	C	D	E	F	G
1	Module Type 20.0kg	SPICN6(LAR)}-60-375/IH | SIE	SPICN6(LAR)}-60-375/IH | NOCT	SPICN6(LAR)-60-380/IH | STC	SPICN6(LAR)-60-380/IH | NOCT	SPICN6(LAR)-60-385/IH | STC	SPICN6(LAR)-60-385/IH | NOCT
2							
3	Maximum Power (Pmax/W)	375	282	380	286	385	290
4							
5	Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
6							
7	Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
8							
9	Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40
10							
11	Short Circuit Current (Isc/A)	11.5	9.23	11.61	9.32	11.72	9.42
12							
13	Module Efficiency	20.90%	20.90%	21.20%	21.20%	21.50%	21.50%

Data extracted after image preprocessing

Chapter 2 - Approach and Results

2.3. Table detection (TD)

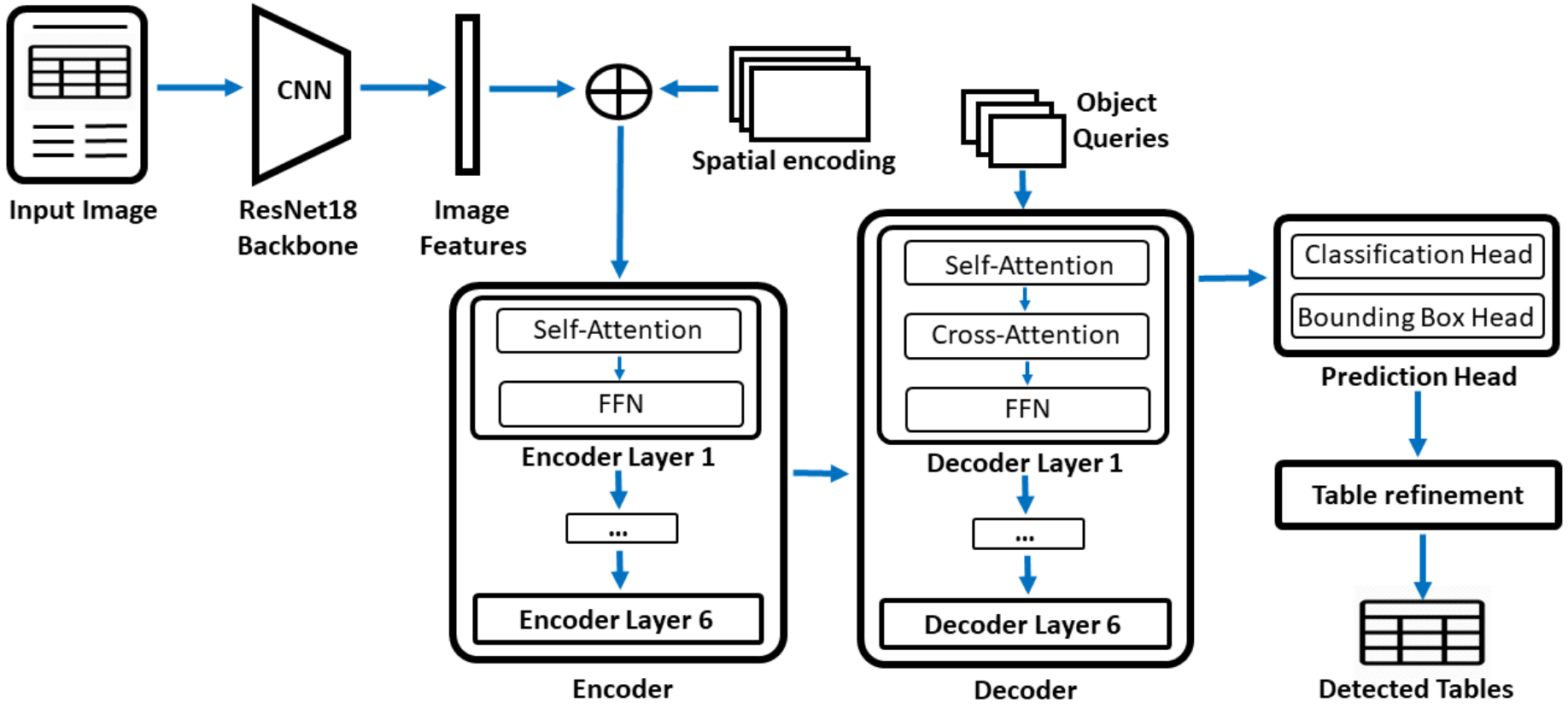


Table Detection architecture diagram

Chapter 2 - Approach and Results

2.3. Table detection (TD) Evaluation Metrics

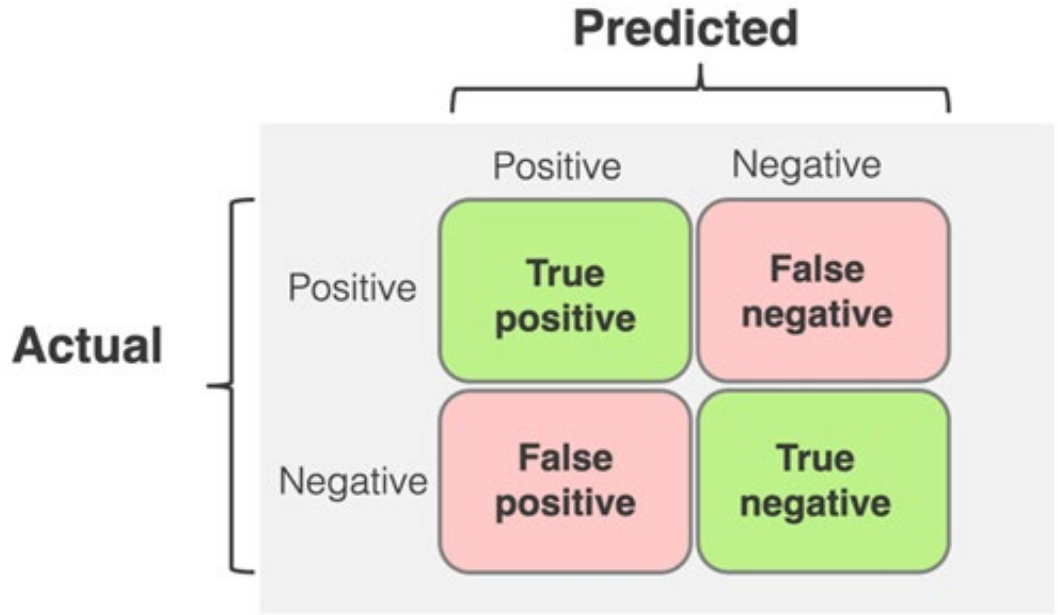
COCO metrics

$$\text{Precision} = \frac{TP}{TP + FP}$$

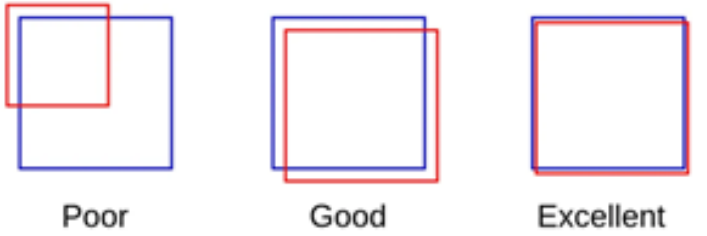
$$\text{Recall} = \frac{TP}{TP + FN}$$

$$\text{F1 Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

$$\text{AP} = \int_0^1 \text{Precision}(r)\, dr$$

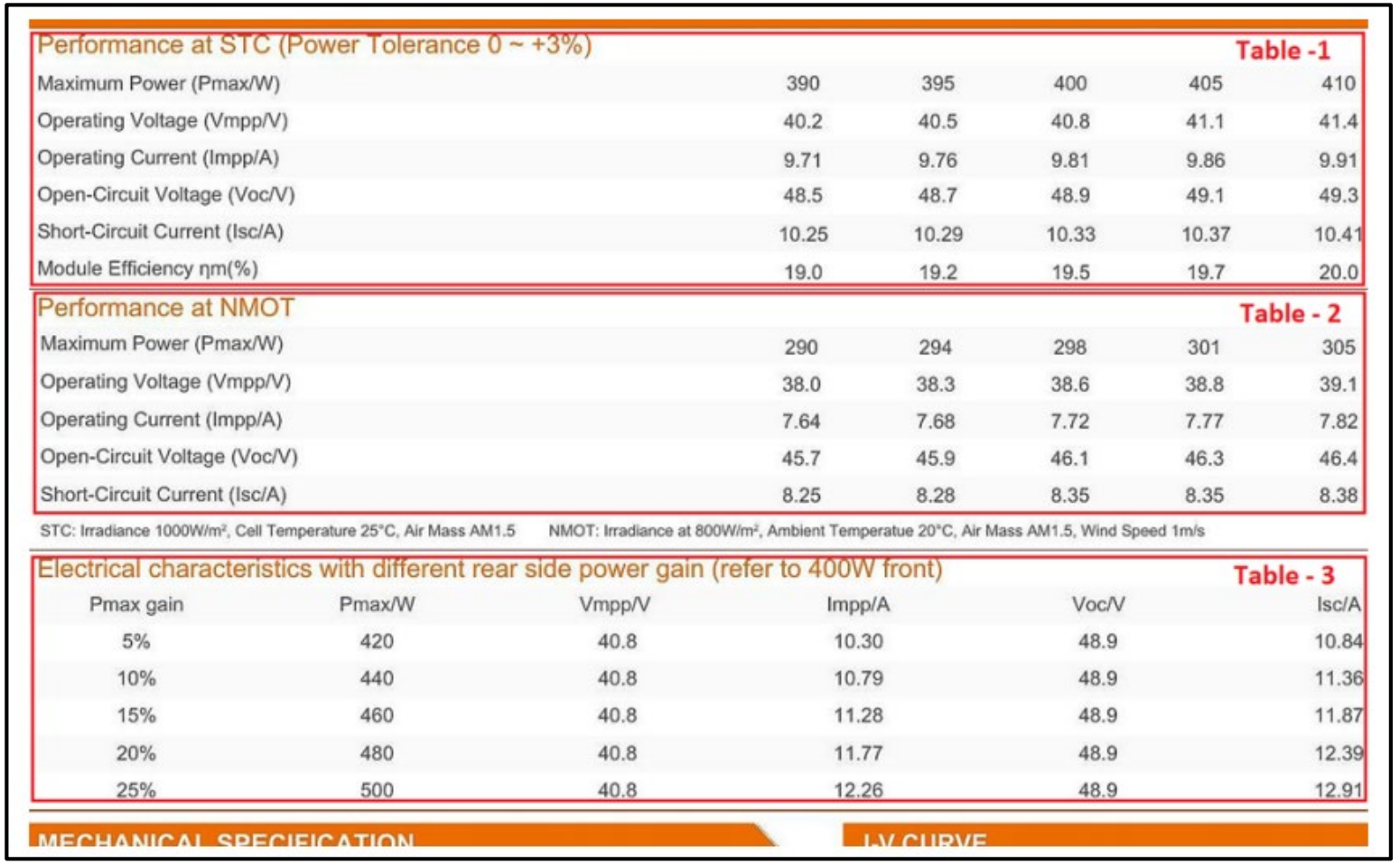


$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



Chapter 2 - Approach and Results

2.3. Table Detection



Performance at STC (Power Tolerance 0 ~ +3%) — Table -1

Maximum Power (Pmax/W)	390	395	400	405	410
Operating Voltage (Vmpp/V)	40.2	40.5	40.8	41.1	41.4
Operating Current (Impp/A)	9.71	9.76	9.81	9.86	9.91
Open-Circuit Voltage (Voc/V)	48.5	48.7	48.9	49.1	49.3
Short-Circuit Current (Isc/A)	10.25	10.29	10.33	10.37	10.41
Module Efficiency ηm(%)	19.0	19.2	19.5	19.7	20.0

Performance at NMOT — Table - 2

Maximum Power (Pmax/W)	290	294	298	301	305
Operating Voltage (Vmpp/V)	38.0	38.3	38.6	38.8	39.1
Operating Current (Impp/A)	7.64	7.68	7.72	7.77	7.82
Open-Circuit Voltage (Voc/V)	45.7	45.9	46.1	46.3	46.4
Short-Circuit Current (Isc/A)	8.25	8.28	8.35	8.35	8.38

STC: Irradiance 1000W/m², Cell Temperature 25°C, Air Mass AM1.5 NMOT: Irradiance at 800W/m², Ambient Temperatue 20°C, Air Mass AM1.5, Wind Speed 1m/s

Electrical characteristics with different rear side power gain (refer to 400W front) — Table - 3

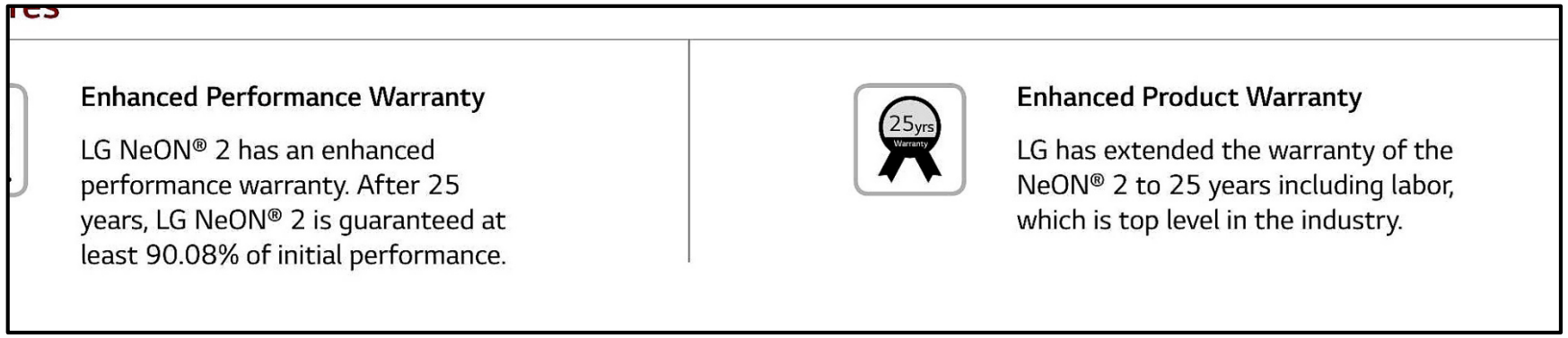
Pmax gain	Pmax/W	Vmpp/V	Impp/A	Voc/V	Isc/A
5%	420	40.8	10.30	48.9	10.84
10%	440	40.8	10.79	48.9	11.36
15%	460	40.8	11.28	48.9	11.87
20%	480	40.8	11.77	48.9	12.39
25%	500	40.8	12.26	48.9	12.91

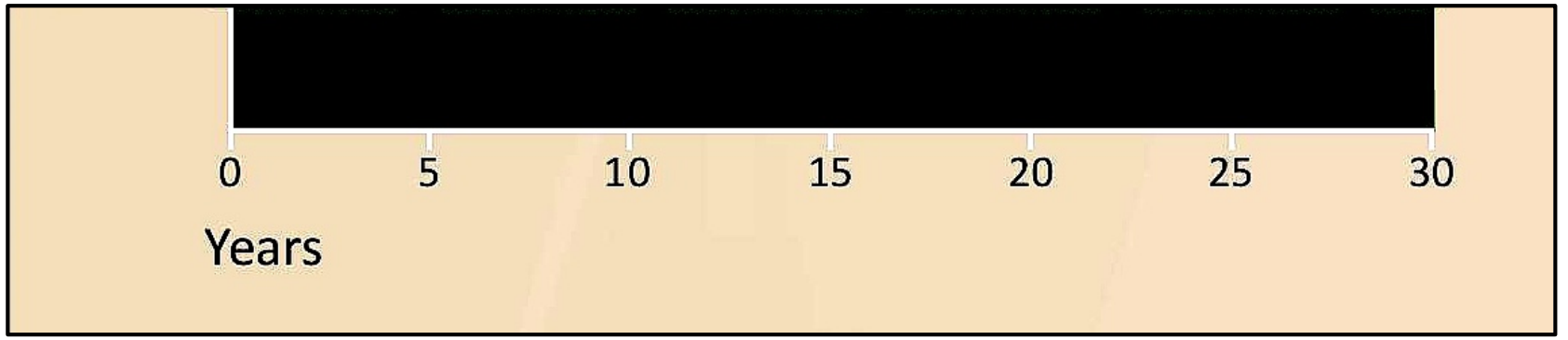
MECHANICAL SPECIFICATION I-V CURVE

Poor detection performance on densely-packed datasheet with missing gridlines

Chapter 2 - Approach and Results

2.3. Table Detection





Well-structured text misidentified as a table

Chapter 2 - Approach and Results

2.4. Table Detection (TD)

Hyperparameter	Value
Learning rate	5×10^{-5}
Batch size	4
Weight decay	1×10^{-4}
LR scheduler	Exponential with Gamma = 0.9
LR drop	1

Hyperparameter configuration of the best table detection model.

Chapter 2 - Approach and Results

2.3. Table Detection

- PubTables-1M dataset - 460,589 annotated document pages
- PV module training dataset – 169 PDF images
- Test dataset – 12 PDF files with15 images

2.4. Table Structure recognition

- PubTables-1M dataset - This dataset has 947,642 fully annotated tables
- PV module training dataset – 545 table mages
- Test dataset – 12 PDF files with 65 table images

Chapter 2 - Approach and Results

2.4. Table Structure Recognition (TSR)

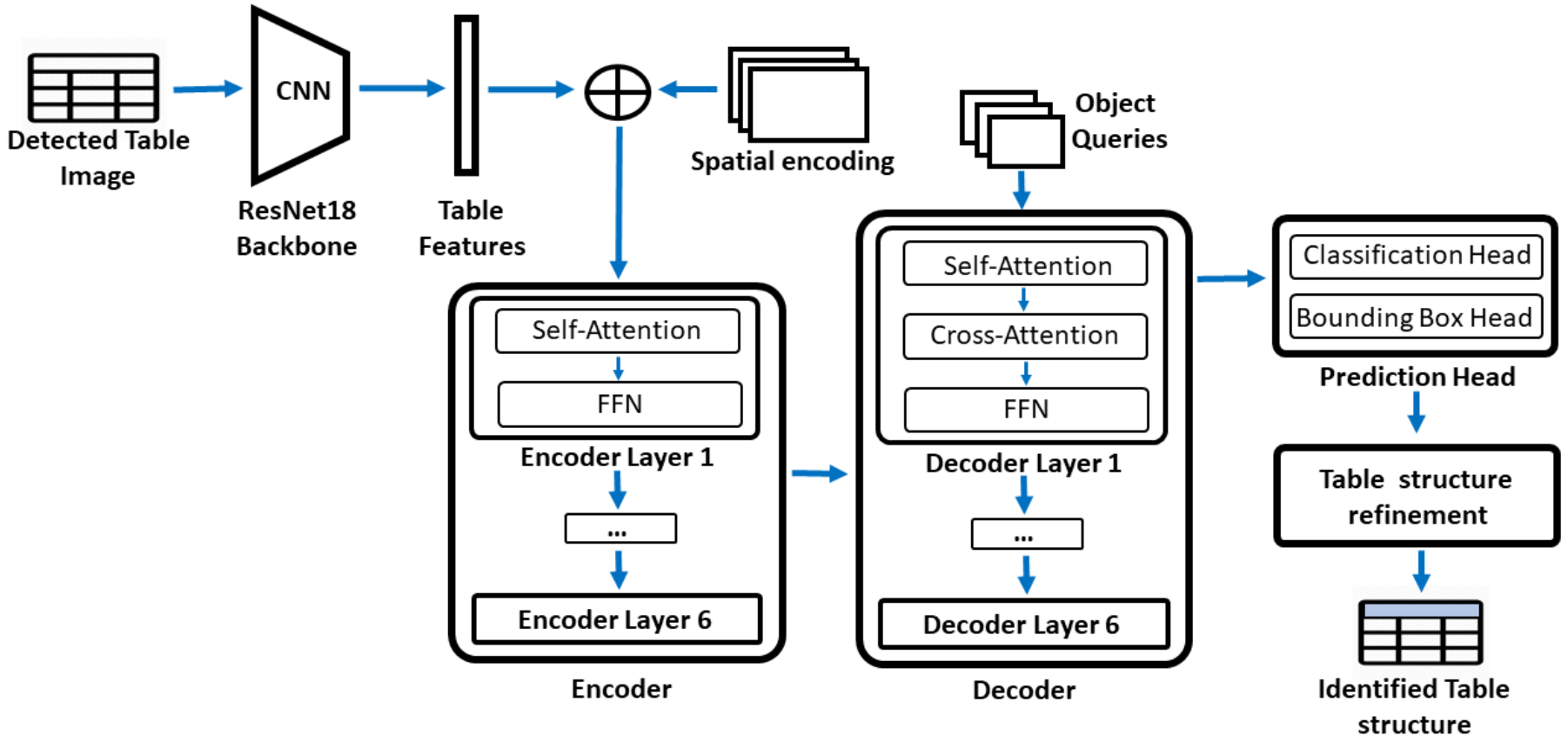


Table Structure Recognition architecture diagram

Chapter 2 – Approach and Results

2.4. Table Structure Recognition (TSR) - Canonicalizaiton Algorithm

	Operation	Purpose
Split blank spanning-cells	Break header-wide blanks into single grid cells.	Prevent giant empty boxes that hide structure
Add missing header rows	If a table row starts with blanks, treat it as an extra column-header row	Capture multi-level column headers.
Expand header rows/cols	Recursively add header rows/cols until every physical column has at least one header cell and vice-versa.	Ensure one-to-one mapping between headers and leaf columns/rows.
Merge identical header spans	Adjacent header cells with the *same* row/column span are merged.	Remove over-segmentation.
Merge contiguous blank cells	Blank cells touching each other inside a header block are combined.	Keep header trees compact.
Project row headers	In dual-axis or horizontal tables, merge all header cells in the projected-row-header row	Support horizontal and two-axis layouts.
Validation	Check that every grid row/column has equal length; drop tables that violate the matrix property.	Guarantee a well-formed output for TDE.

Chapter 2 - Approach and Results

2.4. Evaluation Metrics – Grid Table Sinilarity (GriTS)

- Topological GriTS - Capture cell adjacency and relational structure within a table
- Content GriTS - Associating content with the correct cell structure
- Location GriTS - Assess the accuracy of cell positioning within the table

$$\text{GriTS\_Loc} = \frac{\text{Number of Correct Cell Locations}}{\text{Total Number of Cells}}$$

$$\text{GriTS\_Top} = \frac{\text{Number of Correct Relationships}}{\text{Total Relationships in Ground Truth}}$$

$$\text{GriTS\_Con} = \frac{\text{Number of Correctly Matched Content Cells}}{\text{Total Number of Cells in Ground Truth}}$$

$$\text{GriTS}_f(\mathbf{A}, \mathbf{B}) = \frac{2 \cdot \sum_{i,j} f(\tilde{\mathbf{A}}_{i,j}, \tilde{\mathbf{B}}_{i,j})}{|\mathbf{A}| + |\mathbf{B}|}.$$

$$\text{Accuracy\_Con} = \frac{\text{Number of Correct Cells}}{\text{Total Number of Cells}}$$

Chapter 2 - Approach and Results

2.4. Table Structure Recognition (TSR)

Hyperparameter	Value
Learning rate	1×10^{-4}
Batch size	8
Weight decay	1×10^{-3}
LR scheduler	Cosine Annealing with $T_{\max} = 50$
LR drop	15
Encoder and decoder layers	6

Hyperparameter configuration of the best performing structure recognition model.

Chapter 2 - Approach and Results

2.5. Tabular data extraction

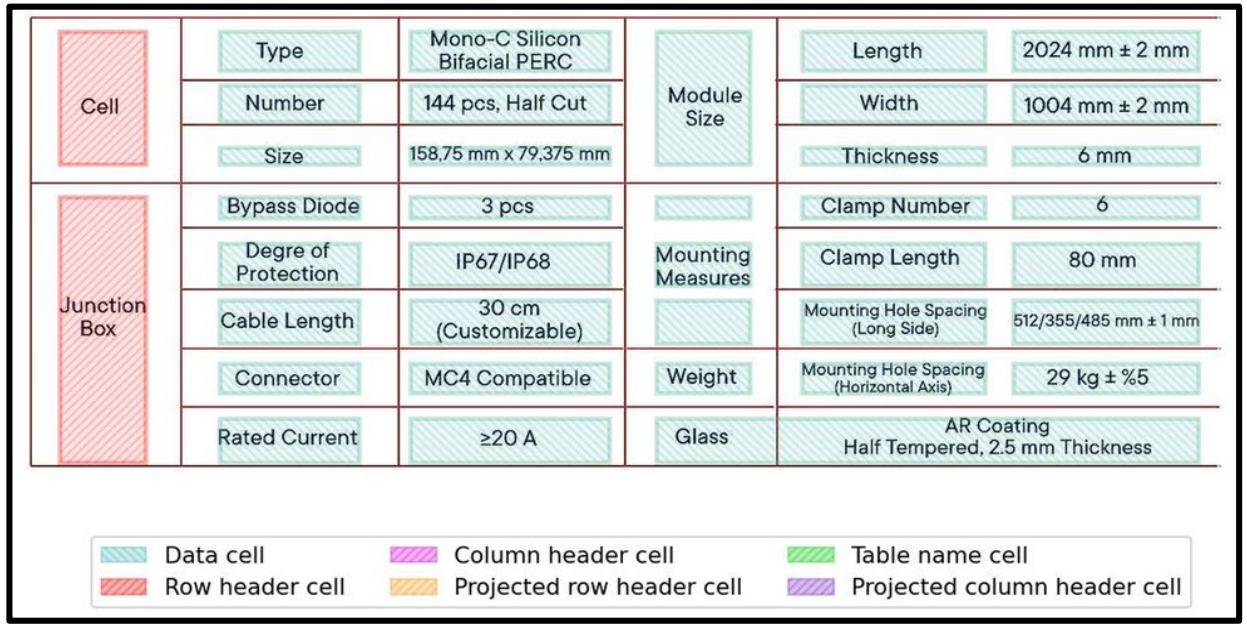
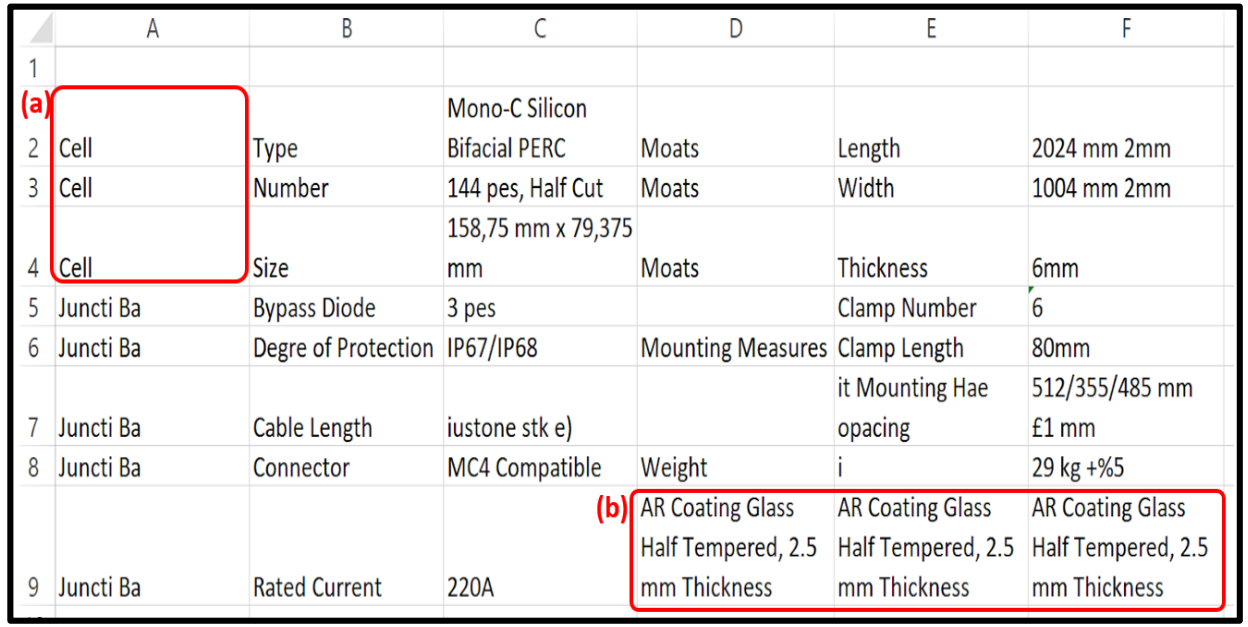


Table structure recognized by the TSR model



	A	B	C	D	E	F
1						
(a)			Mono-C Silicon			
2	Cell	Type	Bifacial PERC	Moats	Length	2024 mm 2mm
3	Cell	Number	144 pes, Half Cut	Moats	Width	1004 mm 2mm
4	Cell	Size	158,75 mm x 79,375 mm	Moats	Thickness	6mm
5	Juncti Ba	Bypass Diode	3 pes		Clamp Number	6
6	Juncti Ba	Degre of Protection	IP67/IP68	Mounting Measures	Clamp Length	80mm
7	Juncti Ba	Cable Length	iustone stk e)		it Mounting Hae opacing	512/355/485 mm £1 mm
8	Juncti Ba	Connector	MC4 Compatible	Weight	i	29 kg +%5
(b) 9	Juncti Ba	Rated Current	220A	AR Coating Glass Half Tempered, 2.5 mm Thickness	AR Coating Glass Half Tempered, 2.5 mm Thickness	AR Coating Glass Half Tempered, 2.5 mm Thickness

Extracted data and exported to Excel

Extraction with significant OCR error and incorrect merged cells captured.

Chapter 2 - Approach and Results

1.3. Lightning-Table vs Table Transformer

Electrical Parameters

Module Type	SPICN6(LAR)-60-375/IH		SPICN6(LAR)-60-380/IH		SPICN6(LAR)-60-385/IH	
	STC	NOCT	STC	NOCT	STC	NOCT
Maximum Power (Pmax/W)	375	282	380	286	385	290
Maximum Power Voltage (Vmpp/V)	35.1	32.5	35.3	32.7	35.5	32.9
Maximum Power Current (Impp/A)	10.69	8.68	10.77	8.75	10.85	8.82
Open Circuit Voltage (Voc/V)	41.6	39.8	41.7	39.9	41.8	40.0
Short Circuit Current (Isc/A)	11.50	9.23	11.61	9.32	11.72	9.42
Module Efficiency	20.9%		21.2%		21.5%	

Table with its internal structure recognized

	A	B	C	D	E	F	G	H	I	J
1	Module Type		SPICN6(LAR)-60- 375/IH			SPIC N6(LAR)-60- 380/IH			SPICN6(LAR)- 60-385/IH	
2										
3		STC		NOCT	STC		NOCT	STC		NOCT
4										
5	Maximum Power (Priax/W)	375		282	380		286	385		290
6										
7	Maximum Power Voltage (Vmpp/V)	35.1		32.5	35.3		32.7	25.5		32.9
8										
9	Maximum Power Current (Impp/A)	10.69		8.68	10.77		875	10.85		8.82
10										
11	Open Circuit, Voltage (Voc/V)	41.6		39.8	417		39.9	41.8		40.0
12										
13	Short Circuit Current (Isc/A)	11.50		9.23	11.61		9.32	172		9.42
14										
15	Module Efficiency		20.9%			21.2%			21.5%	
16										

Data extracted using Lightning-Table

	A	B	C	D	E	F	G
1	Module Type 20.0kg	SPICN6(LAR)}-60-375/IH | SIC	SPICN6(LAR)}-60-375/IH | NOCT	SPICN6(LAR)-60-380/1H | STC	SPICN6(LAR)-60-380/IH | NOCT	SPICN6(LAR)-60-385/IH | S1E	SPICN6(LAR)-60-385/IH | NOCT
2							
3	Maximum Power (Pmax/W)	375	282	380	286	385	290
4							
5	Maximum Power Voltage (Vmpp/V)	Ee|	S25	35:03:00	32.7	35:05:00	329
6							
7	Maximum Power Current (Impp/A)	10.69	8.68	10.77	875	10.85	8.82
8							
9	Open Circuit Voltage (Voc/V)	41.6	39.8 ALT		39.9 A18		40
10							
11	Short Circuit Current (Isc/A)	11.5	9.23	11.61	9.32	11.72	9.42
12							
13	Module Efficiency	20.90%	20.90%	21.20%	21.20%	21.50%	21.50%

Data extracted using Table Transformer

Chapter 2 - Approach and Results

1.3. Lightning-Table vs Table Transformer

Electrical Characteristics					
Power level	435	440	445	450	455
Pmax (W)	435	440	445	450	455
Vmp (V)	41.04	41.24	41.44	41.63	41.82
Imp (A)	10.60	10.67	10.74	10.81	10.88
Voc (V)	49.25	49.44	49.65	49.85	50.06
Isc (A)	11.11	11.17	11.24	11.31	11.38
Module efficiency (%)	20.01	20.24	20.47	20.70	20.93
Maximum system voltage (V)	1500				
Fuse Rating (A)	20				
Temperature coefficient Pmax (%°C)	-0.350				
Temperature coefficient Isc (%°C)	0.05				
Temperature coefficient Voc (%°C)	-0.275				

Table with its internal structure recognized

	A	B	C	D	E	F	G
1	ElectricalCharacteristics						
2	Powerlevel		435	440	445	450	455
3	Pmax （W）		435	440	445	450	455
4	Vmp （V）		41.04	41.24	41.44	41.63	41.82
5	Imp （A）		10.60	10.67	10.74	10.81	10.88
6	Voc （V）		49.25	49.44	49.65	49.85	50.06
7	Isc （A）		11.11	11.17	11.24	11.31	11.38
8	Moduleefficiency （%）		20.01	20.24	20.47	20.70	20.93
9	Maximumsystemvoltage （V）				1500		
10	FuseRating （A）				20		
11	Temperaturecoefficient	Pmax （%°C）			-0.350		
12	Temperaturecoefficient	Isc （%°C）			0.05		
13	Temperaturecoefficient	Voc （%°C）			-0.275		

Data extracted using Lightning-Table

	A	B	C	D	E	F	G
1	Electrical Characteristics	Electrical (	Electrical (	Electrical (	Electrical (	Electrical Characteristics	
2	Power level	435	440	445	450	455	
3	Pmax (W)	435	440	445	450	455	
4	Vmp (V)	41.04	4124	4144	4163	41.82	
5	Imp (A)	10.60	10.67	10.74	10.81	10.88	
6	Voc (V)	49.25	49.44	49.65	49.85	50.06	
7	Isc (A)	11.11	11.17	11.24	11.31	11.38	
8	Module efficiency (%)	20.01	20.24	20.47	20.70	20.93	
9	Maximum system voltage (V)	1500	1500	1500	1500	1500	
10	Fuse Rating (A)	20	20	20	20	20	
11	Temperature coefficient Prax (%°C)	-0.350	-0.350	-0.350	-0.350	-0.350	
12	Temperature coefficient Isc (%°C)	0.05	0.05	0.05	0.05	0.05	
13	Temperature coefficient Voc (%°C)	-0.275	-0.275	-0.275	-0.275	-0.275	

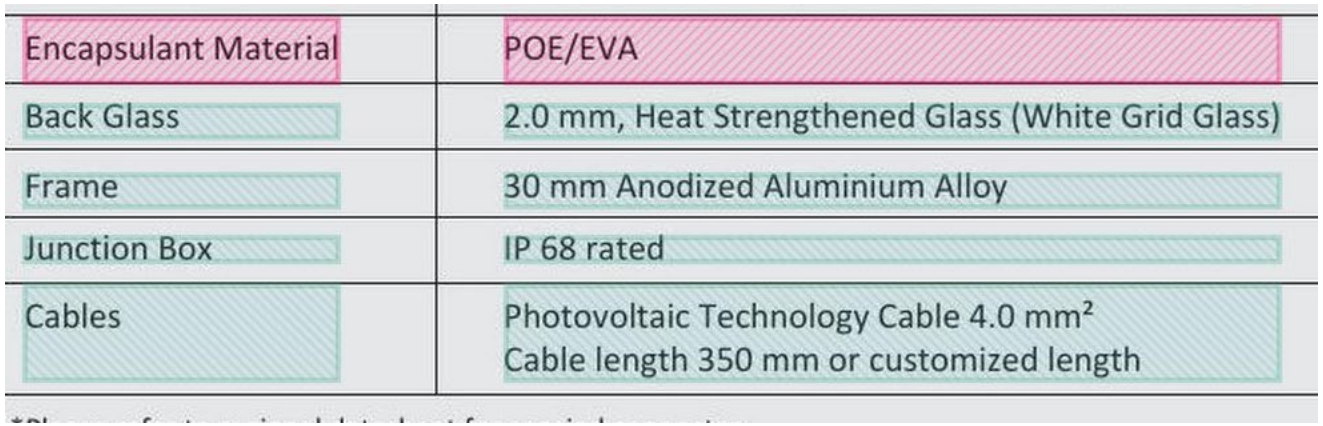
Data extracted using Table Transformer

Chapter 2 - Approach and Results

1.3. Lightning-Table vs Table Transformer

Encapsulant Material	POE/EVA
Back Glass	2.0 mm, Heat Strengthened Glass (White Grid Glass)
Frame	30 mm Anodized Aluminium Alloy
Junction Box	IP 68 rated
Cables	Photovoltaic Technology Cable 4.0 mm²
	Cable length 350 mm or customized length

Data extracted using Lightning-Table



Encapsulant Material	POE/EVA
Back Glass	2.0 mm, Heat Strengthened Glass (White Grid Glass)
Frame	30 mm Anodized Aluminium Alloy
Junction Box	IP 68 rated
Cables	Photovoltaic Technology Cable 4.0 mm² Cable length 350 mm or customized length

Table with its internal structure recognized

5 Encapsulant Material	POE/EVA
Back Glass	2.0 mm, Heat Strengthened Glass (White Grid Glass)
Frame	30 mm Anodized Aluminium Alloy
Junction Box	IP 68 rated
Cables	Photovoltaic Technology Cable 4.0 mm? Cable length 350 mm or customized length

Data extracted using Table Transformer